Effect of Surface Curvature and Chemistry on Protein Stability, Adsorption and Aggregation

Mithun Radhakrishna

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Effect of Surface Curvature and Chemistry on Protein Stability, Adsorption and Aggregation

Mithun Radhakrishna

Enzyme immobilization has been of great industrial importance because of its use in various applications like bio-fuel cells, bio-sensors, drug delivery and bio-catalytic films. Although research on enzyme immobilization dates back to the 1970's, it has been only in the past decade that scientists have started to address the problems involved systematically. Most of the previous works on enzyme immobilization have been retrospective in nature i.e enzymes were immobilized on widely used substrates without a compatibility study between the enzyme and the substrate. Consequently, most of the enzymes lost their activity upon immobilization onto these substrates due to many governing factors like protein-surface and inter-protein interactions. These interactions also play a major role biologically in cell signaling, cell adhesion and inter-protein interactions specifically is believed to be the major cause for neurodegenerative diseases like Alzheimer's and Parkinson's disease. Therefore understanding the role of these forces on proteins is the need of the hour. In my current research, I have mainly focused on two factors a) Surface Curvature b) Surface Chemistry as both of these play a pivotal role in influencing the activity of the enzymes upon immobilization. I study the effect of these factors computationally using a stochastic method known as Monte Carlo simulations.

My research work carried out in the frame work of a Hydrophobic-Polar (HP) lattice model for the protein shows that immobilizing enzymes inside moderately hydrophilic or hydrophobic

pores results in an enhancement of the enzymatic activity compared to that in the bulk. Our results also indicate that there is an optimal value of surface curvature and hydrophobicity/hydrophilicity where this enhancement of enzymatic activity is highest. Further, our results also show that immobilization of enzymes inside hydrophobic pores of optimal sizes are most effective in mitigating protein-aggregation. These results provide us a rationale to understand the role of chaperonins in protein folding and disaggregation. Our results indicate that strong protein-surface interactions and confinement inducement stability inside pores makes it best suitable for enzyme immobilization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Over the last five years the journey which I undertook to understand the ununderstood and to discover the undiscovered has left an incorrigible impression in my mind. I have had the pleasure and privilege of collaborating with some of the brightest individuals in various fields from whom I have learnt a lot. The least I could do is to thank them whole heartedly. Firstly, I want to thank my advisor Prof Sanat Kumar for his guidance and support. It would not be an exaggeration to say that this work would not have been completed without his full cooperation and support. I still vividly remember our first conversation whereupon he told me that the most important thing to take away from a PhD research is the ability to do independent research. I am not sure how far I have succeeded in that, but I am certain that his guidance has shown me the right direction. The progress of an individual during his PhD is analogous to a propagating sinusoidal wave. You encounter crests and troughs at regular intervals. Crests result in publications whereas troughs teach you endurance and I have to specially thank Prof Kumar for pulling me out of those troughs regularly. I also want to thank our collaborator Prof Georges Belfort from RPI,Troy for his guidance and support. I had the honor of coauthoring a manuscript with Prof Belfort's group. Coming from a simulation background, Prof Belfort was instrumental in helping me understand the experimental details and potential applications of my research in industries. I want to thank Prof Angelo Caccuito from the Chemistry department at Columbia University for always making time for me whenever I had questions about simulation methods. I also want to thank Prof Arun Yethiraj from University of Wisconsin, Madison for his guidance on simulations of chemical reactions. I want to thank Dr John Chodera (currently at Memorial Sloan -Kettering center) for all our discussions on WHAM. I had the privilege of taking classes from eminent scientists like

Prof Bruce Berne (Chemistry) and Prof Brian Greene (Physics and Math) who have made learning more enjoyable. I want to thank Joseph Grimaldi from Prof Belfort's group for our collaboration. I want to thank the Department of Energy –Biomolecular materials for funding the project.

I want to thank the current and past group members from the Kumar group. Dan, Thi, Daniel, Mayank, Ellie, Dong, Behnaz, Jagan, Makoto, Babji, Jesse, Michael, Sumit, Joseph, Erica, Shyno, Jestin, Nicolas, Damien, Amber, Manish, Yuping and Hamdi; I have learnt a lot from all of you and thanks for being there whenever I needed help.

I want to whole-heartedly express my sincerest gratitude to Dr Sumit Sharma for teaching me the basics of simulations and guiding me during the first phase of my research. Besides, Sumit has always been there for me as a friend, mentor and as an elder brother. Speaking of mentors, it would be a grave mistake if I don't acknowledge Dr Luis Escobar Ferrand for his constant support and guidance. He has always there to motivate, support and guide me with regard to all aspects of life.

As the proverb goes "All work and no play makes Jack a Dull boy", I want to thank all my friends who provided an opportunity to add social aspect into an otherwise boring and mundane research life☺. I always cherish the dal chawal house dinner hosted by Sumit along with Puneet and Ruttika. I also want to thank Bhuppi, Neha Dobhal, Yanir, Shyno, Joseph, Damla, Kevin, Bala, Arun, Manas, Sonal, Greeshma, Ripla, Parag, Sudha, TJ, K Rao, Aishwarya, Megha, Amrut, Devika, Shubam, Elisa, Arindham, Napoleon, Luis and Shaswath,

Puneet, you know our friendship has been one of the strongest one, I am giving an A+. Sonal, has been my dearest friend with whom even after hours of conversation, I feel like talking more. Joseph has been my strongest tie from the group and also my biking buddy. Greeshma has been my dear friend because of whom I am still not malnourished, thanks for regularly inviting me for dinners. KRao and TJ also helped along the same lines (dinner invitations) during the first two years. Napoleon has been my band-mate. It was so much fun writing music with you pal.

Finally, I want to thank my parents and my sister's family for their constant encouragement and support. I want to specially thank my little niece Anuskha (Anchi) for always cheering me up by her witty and intelligent comments☺.

Dedicated to my parents

Awake Arise Stop not until the goal is reached

Swami Vivekananda

Chapter 1

Introduction

Proteins are macromolecules vital for the biological functioning of living organisms. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses and control growth and differentiation (1-4). Proteins are linear polymers built of monomer units called amino acids, which are linked end to end through peptide bonds. Remarkably proteins spontaneously fold into three dimensional structures that are determined by the sequence of amino acids in the protein structure (5). Proteins contain wide range of functional groups which include thiols, thioethers, carboxylic acids, carboxamides and a variety of basic groups. Depending on the nature of these functional groups amino acids are classified as either hydrophobic or hydrophilic

1.1 Structure

Proteins are made up of 20 basic amino acid residues. Amino acids are made up of amine group ($-NH_2$) and an acid group ($-COOH$). These amino acids undergo condensation polymerization through the removal of water molecule and formation of a peptide bond ($-CO-NH-$) as shown in **Figure 1.1**

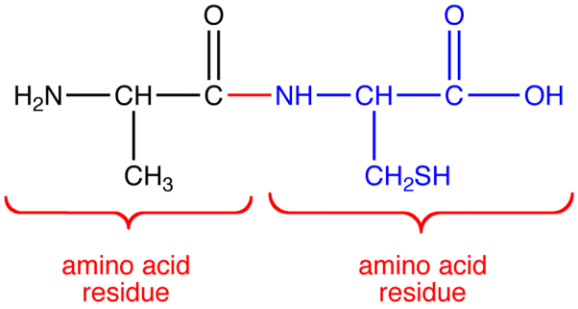


Figure 1.1: Structure of amino acids polymerizing through a peptide bond

Protein structure representation based on hierarchy can be classified into four categories (6-9).

- Primary Structure
- Secondary Structure
- Tertiary Structure
- Quaternary structure

Primary Structure

Proteins are linear polymers formed by linking the α-carbonyl group of amino acid to the α-amino group of another amino acid. The representation of proteins as a linear sequence of its individual amino acid residues is known as the Primary structure. This arrangement of amino-acid residue is what that drives proteins to fold into secondary structures. Hence by looking at the primary structure we can infer upon the α-helix or β-sheet forming tendency of the protein.

Secondary Structure

Polypeptides fold into unique three dimensional structures like α-helices, β-sheets, β turns under the influence of hydrogen bonding between the amino acid residues. Individual domains of most of the amino acids are identified to be α-helices or β-sheets (**Figure 1.2**)

Tertiary Structure

It is the most complex structure of a protein formed by the self assembly of numerous secondary structures like α-helices, β-sheets and random coils. In fact, most of the proteins in their active state remain as tertiary structures (**Figure 1.2**)

a) **b)** **c)**

Figure 1.2: Different levels of protein structure a) α-helix b) β-sheets c) Tertiary structure

Protein surface interactions play a vital role in modulating cell adhesion, triggering the biological cascade resulting in foreign body response, central to the diagnostic assay/sensor device and also initiate bio-adhesion like marine fouling and bacterial adhesion (10-14).

1.2 Protein Adsorption

Protein adsorption is a major consequence of protein surface interactions. Exposure of an aqueous protein solution to a solid surface typically results in an excess accumulation of protein molecules at the solid/liquid interface. Adsorption behavior of proteins on surfaces is the net result of various types of interactions. Proteins often tend to lose their native structure upon adsorption due to the structural rearrangement of proteins on the surface. In some proteins which are structurally very coherent, this structural rearrangement is minimal and adsorption of such proteins is mainly driven by electrostatic forces (15, 16). Hydrophobic effect is also one of the

major driving forces behind protein adsorption. The contribution of dehydration of a hydrophobic surface to the overall Gibbs free energy of free adsorption is believed to be greater than the contribution from electrostatic attraction (15, 17, 18).

Many experiments have reported on the structural changes the protein undergoes upon adsorption onto variety of surfaces. One of the most reported observation is the reduction in the α-helix content and increase in the β-sheet and random coil content of the protein upon adsorption (19-23). They are few exceptions to the above generality. Zoungrana et al reported an increased α-helix content in the protein *α-chymotrypsin* upon adsorption onto a hydrophobic surface of Teflon whereas the adsorption of the protein *cutinase* resulted in decreased helical content (24). Proteins upon adsorption expand and gain conformational entropy compared to the α-helix state. Further expansion provides more surface area for protein-surface interactions which also contributes to the decrease of free energy of the system. Structural rearrangements can also be facilitated through inter-protein interactions between the adsorbed proteins. Sethuraman et al (25, 26) showed that at high protein concentrations denaturation of proteins upon adsorption was mainly due to inter-protein interactions.

1.3 Thermodynamics of protein adsorption

All equilibrium processes are thermodynamically driven accompanied with the decrease in free energy. The free energy is made up of two components energy and entropy. The change in free energy upon adsorption is given by $\Delta G_{ads} = \Delta H_{ads} - T\Delta S_{ads}$, where ΔH_{ads}, ΔS_{ads} represent the enthalpic and entropic change upon adsorption and T is the thermodynamic temperature. Unlike adsorption of colloids protein adsorption is accompanied with structural rearrangements on the

surface. Proteins in their active state fold into secondary structures and self assemble to form domains. Upon adsorption these compact domains begin to unfold and start expanding on the surface. Protein unfolding on the surface contributes to a gain in the conformation entropy accompanied with the favorable enthalpic contribution due to protein-surface interaction. Further, adsorption on hydrophobic surface also involves expulsion of ordered water molecules near the hydrophobic surface which results in an increased entropy of the water molecules (22, 27, 28). But, structural rearrangements also involve breaking of some of the internal hydrogen-bonds which is energetically not favorable. Hence, this delicate balance between energetic loss and entropic gain governs protein adsorption on surfaces.

My research work mostly focuses on designing/modifying surface for enzyme immobilization with a goal to retain/enhance enzymatic activity upon immobilization. Hence understanding protein-surface interactions form the crux of this study. Many experimental works show increased enzymatic activity when enzymes are immobilized inside mesoporous materials (29-33). This has encouraged us to investigate the stability of proteins under confinement.

1.4 Confinement effects on structural stability of protein

The cellular environment in which a protein folds and performs its functions is crowded with several biological molecules including lipids, carbohydrates, and other proteins. Most of the experimental, theoretical, and computational studies on protein folding however have relied on studying proteins in the infinitely dilute limit. This idealized dilute environment is different from that, inside the cell, even if the specific interactions between the protein and the surroundings are minimal. Geometrical restrictions imposed by the neighboring molecules can have an appreciable impact on protein structure, merely by virtue of their excluded volume. Nature, in

fact utilizes such phenomena to its advantage. A protein inside a confined space will be stabilized by folding forces different from those for proteins in bulk solutions, that is away from any confining walls. In particular, some expanded configurations of the unfolded chain will not be allowed inside the confined space, due to excluded volume. Examples of such spaces include the pores within chromatographic columns, the Anfinsen cage in chaperonins, the interiors of ribosome, or regions of steric occlusion inside cells As a result of the aforementioned oberservations, folding under confinement is emerging as an active area of research. Several recent studies including experimental (32, 34, 35), theoretical (36-39) and computational work (40-42) have been conducted to understand this phenomenon. The purpose of present work is to identify the behavior of proteins when a confined boundary interacts with the protein under different potentials and study the effect of theses potentials on protein stability. **Figure 1.3** is reproduced from the work by Dill et al (36) on the effect of athermal surfaces on polymers.

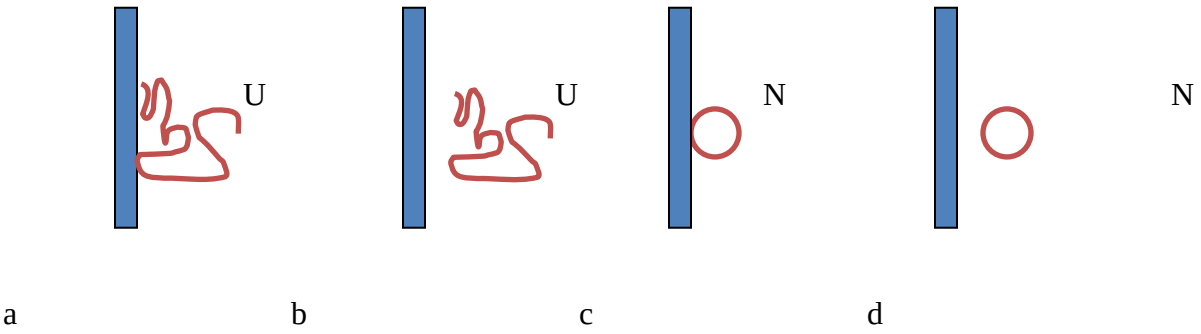


Figure 1.3: Effect of a wall or boundary on the configurations of a protein. In the unfolded state U, configurations (a) are not viable since the protein pass across the steric boundary. Configurations (b) is however viable. In the native state N, the protein is modeled as a sphere, so there is only a single configuration. The native protein cannot be located in position c but can be located in position d.

1.5 Enzyme immobilization

The process of tethering enzymes onto surfaces physically/chemically for their function as catalysts is described as enzyme immobilization. Immobilizing enzymes increases the efficiently of the process simultaneously making the enzyme reusable. Enzymes exhibit high activities and selectivity under moderate reaction conditions and hence have attracted attention as catalysts in the field of green and sustainable chemistry. Immobilization of the enzyme onto solid supports such as polymers, ceramics and porous silicas provide an attractive method to overcome the problems caused due to the stability and reusability issue associated with it (31, 43-46). Many previous works have focused on immobilizing enzymes onto surfaces of nanoparticles and other nano-structures but have suffered due to loss in enzymatic activity because of both inter-protein and protein surface interactions. Many experimental (31, 43) and theoretical (47-49) studies have reported increased thermal and structural stability of proteins under crowded and confined environments. It is believed that under confinement the unfolded state loses more entropy compared to the folded state and consequently the folded (active) state of the protein is stabilized. Based on the above observations enzyme immobilized inside porous media should exhibit higher stability and activity. Therefore mesorporous silica has become a viable candidate as a substrate to immobilize enzymes. This immobilization can be achieved through physical adsorption, covalent attachment of the enzyme onto the pore walls, encapsulation or trapping the enzyme inside pore channels or formation of cross-linked cage like structures to trap the enzymes (50). Although it is believed that physical adsorption of the proteins onto the pore walls is simple and cost effective it suffers from the fact that weak interactions between the surface and the enzyme is not enough to keep the enzyme immobilized. Encapsulation is the physical

confinement of the guest enzyme into the host support matrix. The matching of support materials and enzyme size must be carefully considered. Enzymes with sizes equal to or larger than that of the host system will have lower loadings and simply adsorb on the external while those with sizes much smaller than the host material will also freely leach from the within the host. Covalent attachment involves binding amino –acid residues of the enzyme to the support matrix. This method is popular for high surface area support matrixes with large pore diameters where substrate and product can freely diffuse without the worry of enzyme leaching. Unfortunately, the covalent attachment of enzymes to support generally lowers activity of the enzymes and efforts are being made to design and modify the surface to make it compatible for enzyme immobilization.

1.6 Protein Model

The philosophy underlying the model is as follows. We seek a model that is based on the dominant physical driving forces – the hydrophobic interactions, conformational free excluded volume of the chain, and the steric restrictions imposed by the excluded volume of the chain. We seek a model without adjustable parameters or arbitrary approximations, from which we can draw rigorous inferences and with which we can determine the native states for all the different sequences of monomers.

Protein chains are modeled as copolymers of specific sequences of hydrophobic (H) and polar/charged (P) monomers (HP model). Chains are configured as self avoiding flights on three dimensional simple cubic lattices. A monomer can make at most 6 contacts with nearest neighbor sites on the simple cubic lattice. A contact between two H monomers has a favorable

free energy and the contact free energy for all other types is 0. Conformations are native if they have the lowest free energy i.e. the maximum number of HH contacts over all possible conformations. While the HP lattice model is crude, it has the following characteristics of real proteins; when the HH attraction is strong, chains fold to a few native states with non polar cores and with secondary structures defined by topological contacts. The HP model also resembles real proteins in some mutational and kinetic properties. Most importantly the model has the same conformational search problem that proteins have. That is, both proteins and the HP lattice model have the same conformational spaces too large to be enumerated by a computer (51-53).

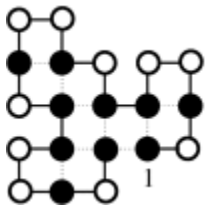


Figure 1.4: 2D HP model of 20mer protein. Filled circles H groups and unfilled circles P groups

Chapter 2

Simulation Methods

2.1 Monte Carlo Simulations

Monte Carlo (MC) simulation is a stochastic technique used in many fields like chemistry, biology, engineering and finance to understand the behavior of complex systems. The simulation technique is used to study the equilibrium behavior given a probability distribution to model it (54-57). In my research, I have used MC to characterize the stability of proteins using coarse grained protein models Below, I discuss some of the important Monte Carlo techniques used in my simulation study. Gaussian distribution is most commonly used to model the behavior of most of the equilibrium systems. (58-60)The Gaussian probability density function of a random variable X with a mean μ and standard deviation σ is given by

$$P_X(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

2.1.1 Metropolis Sampling

Metropolis sampling is the most widely used Monte Carlo technique used to compute thermodynamic averages. Originally developed by Metropolis et al to simulate the freezing transition of hard spheres, today it has been used to estimate the values of multidimensional integrals in any context they arise (61). The core of the Metropolis Monte Carlo algorithm is in the way it selects samples from a known probability distribution function which is Gaussian in my study.

In the literature of numerical analysis, Metropolis sampling is classified as an importance sampling technique. Importance sampling increases the rate of convergence of the algorithm as it picks the sample from a desired probability distribution rather than from a uniform distribution (62-64). In this algorithm successive states of the system are generated to generate a special type of random walk known as the Markov chain. A Markov chain is a random walk in the density space where the conditional probability of evolution from a state depends solely on the present state and is independent of the prior states visited (65). Mathematically a Markov process can be represented as in equation (2)

$$P(x_n = a \mid x_{n-1} = b, x_{n-2} = c, ..., x_o = 0) = P(x_n = a \mid x_{n-1} = b) \quad (2)$$

Given a equilibrium probability distribution as a function of the reaction coordinate q; ρ(q), let C($q_i \rightarrow q_j$) be the conditional probability of transition from state i to state j. Then the probability of transition from state i to j is given by

$$P(q_i \rightarrow q_j) = C(q_i \rightarrow q_j)\rho(q_i) \quad (3)$$

Inorder to approach a steady state at equilibrium the transitional probability should satisfy the fact that on an average the number of transitions from state i to j is same the number of transitions in the reverse direction. This is one of the most important requirements for the MC technique to work and is known as Ergodicity (66, 67). Random walks whose steady state distribution does not depend on the initial state/configuration are known to execute an ergodic walk in the configuration space. The reversibility condition of the random walk is known as detailed balance which can be represented mathematically as shown in equation 4.

$$C(q_i \to q_j)\rho(q_i) = C(q_j \to q_i)\rho(q_j) \quad (4)$$

If the algorithm satisfies the detailed balance condition as outlined above it asymptotically approaches the equilibrium probability distribution according to ρ(q). The transition rate can be written in terms of probability of acceptance (A) and probability of selecting a particular state (ω) as shown in equation 5

$$C(q_i \to q_j) = A(q_i \to q_j)\omega(q_i \to q_j) \quad (5)$$

To satisfy the detailed balance condition we have

$$\frac{A(q_i \to q_j)}{A(q_j \to q_i)} = \frac{\omega(q_j \to q_i)\rho(q_j)}{\omega(q_i \to q_j)\rho(q_i)} = r \quad (6)$$

The value of r is estimated from equation (6) and is compared with a real random variable generated from a uniform distribution in the interval (0,1) and finally accepted with a probability min(1,r). This random walk is continued until the thermodynamic averages converge to the equilibrium value at that given condition. This is the general principle behind Metropolis sampling.

2.1.2 Canonical Ensemble Simulations (NVT)

Inorder to carry out molecular simulations, few parameters are held fixed for a given system and the quantity of interest which is the observable is tracked as a function of time. In a canonical ensemble the number of atoms (N), Volume (V) and the thermodynamic temperature (T) are held constant. The equilibrium distribution of a canonical ensemble ρ(q) and the thermodynamic

average of a observable f(q) is given by equations 7 and 8 respectively. β^{-1} is the thermal fluctuation energy (k_BT) , U(q) is the potential energy associated with a configuration q.

$$\rho(q) = \frac{\exp(-\beta U(q))}{\int dq \exp(-\beta U(q))} \tag{7}$$

$$< f >= \frac{\int dq f(q) \exp(-\beta U(q))}{\int dq \exp(-\beta U(q))} \tag{8}$$

$$Z = \int dq \exp(-\beta U(q)) \tag{9}$$

k_B is the Boltzmann constant, Z is called the partition function and is related to the free energy (F) by the equation $F = -k_B T \ln Z$ .The acceptance criterion for canonical ensemble simulations is given by $\min(1, \exp(-\beta \Delta U))$.

2.1.3 Isobaric Isothermal Ensemble Simulations (NPT)

In a NPT ensemble the number of molecules (N), Pressure (P) and Temperature (T) are held constant and the volume (V) is allowed to vary. Since most of the real world experiments are carried out at constant Temperature and Pressure NPT Monte Carlo is a widely used simulation technique. NPT simulation was first carried out by Wood et al in 1957 (68) for the simulation of hard sphere disks. The acceptance criterion for a NPT MC simulation is given by

$$\min(1, \exp\{-\beta[U(q_1, V_1) - U(q_{2,} V_2) + P(V_1 - V_2) - (N+1)\beta^{-1} \ln(V_1 / V_2)]\}) \tag{10}$$

2.1.4 Grand Canonical Ensemble (μVT)

In this ensemble the chemical potential (μ), Volume (V) and Temperature (T) are held constant and the number of atoms (N) is allowed to change. Grand canonical (GC) ensemble serves as an indispensible tool to study phenomenon like adsorption (69-71). The acceptance criterion for a GCMC simulation is given by

$$acc(N \to N+1) = \min[1, \frac{N}{V} \exp(-\beta\{\mu + U(N-1) - U(N)]\}\} \quad (11)$$

2.2 Trial Moves

The trial move forms the integral part of the Monte Carlo technique. Depending on the system studied, various moves like displacement, rotation, chain regrowth, reptation are employed. For the most part of my study, I deal with lattice polymers. Standard moves like internal flip, end bond flip and crankshaft moves were employed (72-74). They are depicted in **Figure 2.1**

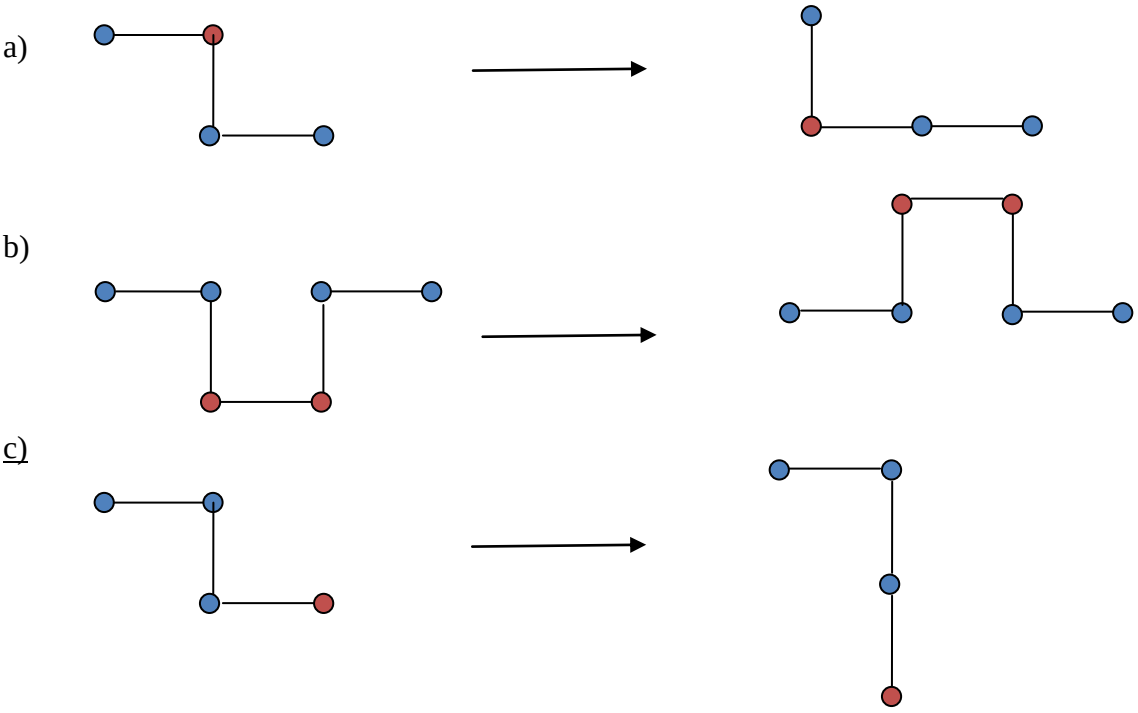


Figure 2.1: Monte Carlo move sets for lattice polymers; a) Internal flip b) Crankshaft flip c) End bond flip

2.3 Configuration Bias Monte Carlo (CBMC)

For simulating systems near melting temperatures, conventional lattice moves as described above are not sufficient to sample the phase space due to large free energy barriers. Therefore sophisticated moves like Rosenbluth moves and CBMC are required. The chain growth is biased towards low energy conformations and this bias is removed during the acceptance criterion. This involves randomly selecting a segment *i* from a chain of *N* segments and subsequently removing either segments *i* to *N* or *1* to *i*, with equal probability and then re-growing the chain with a bias towards lower energy conformations. This bias is removed during acceptance. A segment *i*, can be grown in *k* different directions. The energy of a trial direction *j* is denoted by $u_i(j)$. From the *k* possible directions, we select one, say *n*, with a probability $p_i(n) = \frac{\exp(-\beta u_i(n))}{w_i(n)}$, where $w_i(n)$ is defined as $w_i(n) = \sum_{j=1}^{k} \exp(-\beta u_i(j))$. This step is repeated until the entire chain is grown and the Rosenbluth factor *W(new)* for the new chain is calculated. *W(new)* is defined as *W(new)*= $\prod_{i}^{N} w_i(n)$. The old chain is retraced and *W(old)* is calculated. This chain regrowth move is accepted with a probability $\min(1, \frac{W(new)}{W(old)})$.

2.4 Weighted Histogram Analysis Method (WHAM)

The thermodynamic average of an observable, A, which is a function of U (U is the internal energy of the system under study) at any given temperature can be calculated mathematically using Eq 12, where A(U) is defined as the average of A over all configurations with an energy U

and Ω (U) - the density of states (DOS) - represents the number of different configurations which have an energy U.

$$\langle A \rangle_T = \frac{\int dU \Omega(U) \exp(-U / k_B T) A(U)}{\int dU \Omega(U) \exp(-U / k_B T)} \tag{12}$$

Ω(U) is independent of temperature and is sufficient to calculate the thermodynamic averages of the system at any given temperature. This requires calculation of Ω(U) for all possible values of U through thorough sampling of phase space. Since the free energy landscape of proteins is very rugged due to the existence of many metastable states, canonical Monte Carlo simulations were carried out at various temperatures (T^*) and surface hydrophobicity (λ) so as to thoroughly explore phase space. These simulations give us the local free energy landscape of the system. All the data from these simulations can be combined using the WHAM to get the global free energy landscape. The DOS and free energy of the system are estimated directly from WHAM using a self consistent iterative process represented by Eqs 13 and 14.

$$\Omega_m = \frac{\sum_{k=1}^{K} g_k^{-1} H_{mk}}{\sum_{k=1}^{K} g_k^{-1} N_k \Delta U \exp\left[f_k - \beta_k U_m \right]} \tag{13}$$

$$f_k = -\ln \sum_{m=1}^{M} \Omega_m \Delta U \exp(-\beta_k U_m) \tag{14}$$

Here, K is the total number of simulations, $\boldsymbol{\Omega_m}$ is the DOS of the state *m*, $\boldsymbol{f_k}$ is the free energy of the k^{th} simulation, and $\boldsymbol{g_k}$ is the statistical uncertainty associated with k^{th} simulation. $\mathbf{g_k = 1 + 2 * \tau}$, where $\boldsymbol{\tau}$ is the area under energy autocorrelation curve**.** $\boldsymbol{H_{mk}}$ is the histogram entry of the m^{th}

state from the k^{th} simulation. $\mathbf{N_k}$ is the total number of histogram entries from the k^{th} simulation. $\boldsymbol{\beta_k}$ is the inverse of the thermodynamic temperature ($=\frac{1}{k_B T}$) of the k^{th} simulation. U_m is the energy associated with the m^{th} state and $\boldsymbol{\Delta}U$ is the width of the histogram bins, which is 1 in our current study. Eqs. 13 and 14 are solved iteratively for self-consistency to estimate the DOS and free energy of the system (75).

It should be noted that the exclusion of g_k or incorrect estimates of g_k result in an inaccurate estimation of the DOS and free energy. The effect of g_k manifests itself prominently for simulations carried out near the folding temperature of the protein. At these temperatures, the system takes a very long time to relax and hence simulations should be carried out considerably longer than the appropriate system relaxation times to get the correct estimate of g_k.. The instantaneous normalized energy auto correlation function is given by

$$C(t)=\sum_{\tau=1}^{N}\frac{\left(U(t)-\langle U\rangle\right)-\left(U(t+\tau)-\langle U\rangle\right)}{\left(U(t)-\langle U\rangle\right)\left(U(t)-\langle U\rangle\right)}$$

Near the folding temperature of the protein, it is very hard to overcome the free energy barriers since the Boltzmann acceptance probability at these temperatures is small. As a consequence, simulations carried out for short times underestimate the value of C(t), which would result in large errors in the estimation of the DOS of the system.

2.5 Parallel Tempering Simulations: As discussed above, simulations of large polymers like Proteins, Lipids and other biological molecules is very difficult at low temperatures since various competing interactions yield to frustration and a rough energy landscape. Hence at low

temperatures the simulations get entrapped in a multitude of local minima separated by high energy barriers.

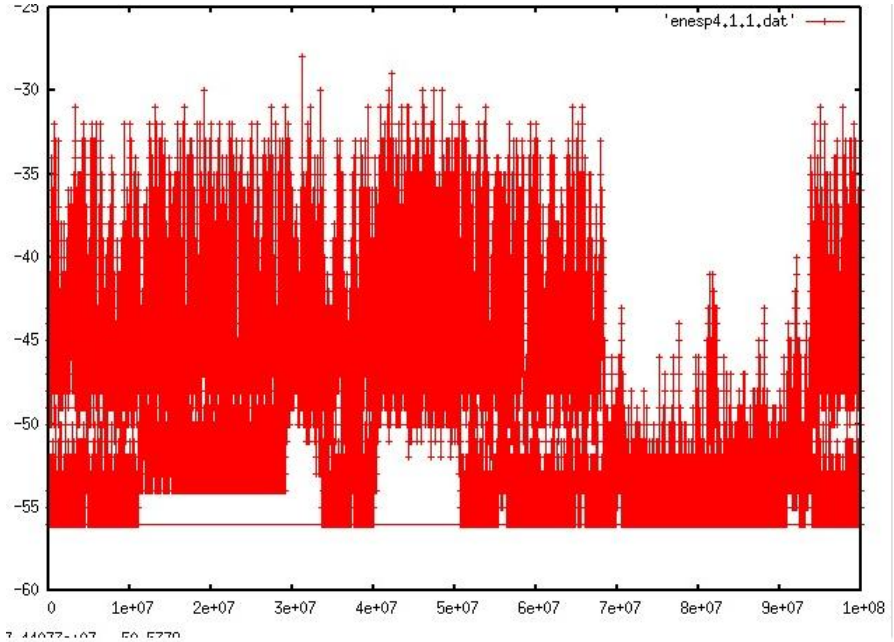


Figure 2.2: Plot of the energy of the 64mer HP protein model as a function of time from a Parallel Tempering simulation.

Parallel tempering technique aims to alleviate this problem by running simulations around each of these local minima in parallel and subsequently swapping the adjacent systems. As a result, systems in every local minima explore all other minima until a global minimum is found and thence a thorough sampling of the entire phase space is obtained (76-78).

In a parallel tempering algorithm one considers an artificial system built up of N non-interacting copies of the same system but each at different temperatures ($k_B T = \frac{1}{\beta}$). The total states of the systems are described by the vector $C = (c_1, c_2, c_3, ..., c_n)$ each of the system at different

temperatures $\beta = (\beta_1, \beta_2, \beta_3, ..., \beta_n)$. Let's assume β1 < β2 <….<βn in general and in order to approach the Ergodic behavior of the chains discussed before, we have to construct moves such that it preserves the Markovian chain property. This is can be achieved by running standard MC

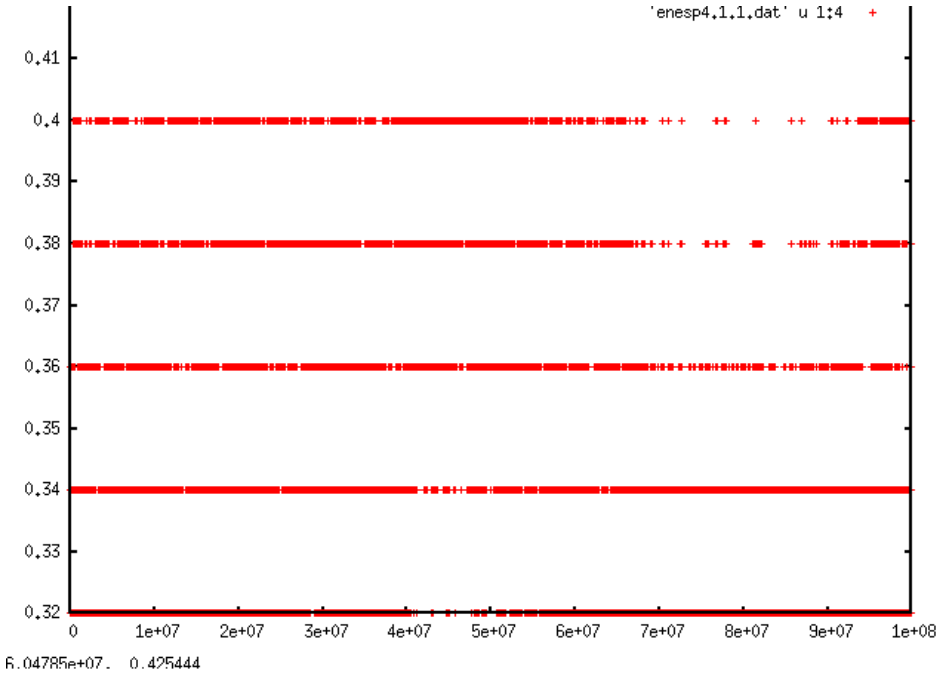


Figure 2.3: Temperature distribution of a replica in Parallel Tempering simulation of the 64mer HP protein as a function of time. We see that the simulation explores all the temperatures uniformly (Temperature range T*=0.32-0.40)

simulations in each of the N systems at their respective thermodynamic temperatures with the combination of swapping the coordinates of the system at regular intervals. Inorder to accomplish the swapping moves, two adjacent systems are randomly selected and then the coordinates are swapped with a acceptance probability $\min(1, \exp(\Delta\beta\Delta E))$, where $\Delta\beta = (\beta_2 - \beta_1)$ and $\Delta E = (E_2 - E_1)$. This ensures proper sampling of the all the local minima. At equilibrium, it can be verified that the thermodynamic average of the observable in all the N systems should be around the same value. The temperature and energy from a single replica is

depicted in **Figure 2.2** and **Figure 2.3** respectively. Simulations should be carried out until the each replica has sampled the entire temperature window uniformly as shown in **Figure 2.3.**

2.6 Wang Landau sampling

Monte Carlo simulations give us the thermodynamic averages of the observables through a stochastic simulation which follow a Boltzmann distribution model. One of the major challenges with the metropolis sampling technique is the fact that at temperatures near folding/melting the system takes a long time to relax and as a result thorough sampling of the phase space is not possible. The Boltzmann probability is proportional to $\exp(-H/k_BT)$ where H is the Hamiltonian of the system. Evidently, at low temperatures the acceptance probability is very low and as a result it becomes difficult for the system to cross the free energy barriers and consequently results in inaccurate estimates of the thermodynamic averages.

To circumvent this dependence of sampling on temperature (T), Wang and Landau came up with a novel non-Boltzmann sampling technique where the probability of finding the system in the state q is instead proportional to $\frac{1}{\Omega(q)}$ (79-83), where Ω(q) is the density of states (DOS) of the system in the state q. The core concept lies in the fact that since DOS of the system is unknown apriori, we start with any real value of Ω and subsequently sample the system according to the non-Boltzmann probability as stated above to get a flat histogram. The process is continued until the values of Ω converge. The acceptance probability for a transition from the state q_1 to the state q_2 is given by $\min(1, \frac{\Omega(q_1)}{\Omega(q_2)})$. A detailed explanation on the WL algorithm and other methods incorporated to increase the efficiency of the WL algorithm is given in Chapter 5.

Chapter 3

Stability of proteins inside a hydrophobic cavity

We study the effects of confinement and hydrophobicity of a spherical cavity on the structural and thermal stability of proteins in the framework of a Hydrophobic-Polar (HP) lattice model. We observe that a neutral confinement stabilizes the folded state of the protein by eliminating many of the open chain conformations of the unfolded state. Hydrophobic confinement always destabilizes the protein due to protein-surface interactions. However, for moderate surface hydrophobicities, the protein remains stabilized relative to its state in free solution because of the dominance of entropic effects. These results are consistent with our experimental findings of (a) enhanced activity of alcohol dehydrogenase (ADH) when immobilized inside the essentially cylindrical pores of hydrophilic mesoporous silica (SBA-15) and (b) unaffected activity when immobilized inside weakly hydrophobic pores of methacrylate resin compared to its activity in free solution. In the same vein, our predictions are also consistent with the behavior of Lysozyme and Myoglobin in hydrophilic and hydrophobic SBA-15, which show qualitatively the same trends. Apparently, our results have validity across these very different enzymes, and we therefore suggest that confinement can be used to selectively improve enzyme performance.

3.1 Introduction

The presence of proteins in confined and crowded environments is common in biological systems. Cellular environments, the Anfinsen cage in Chaperonin, the interior of ribosomes and the pores in chromatographic columns are a few examples where such crowding effects are routinely encountered (84-88). It has also been reported that the fractional volume of crowding molecules inside the cytoplasm can be as high as 20-30% (89-91). Under such conditions, the

likelihood of proteins aggregating or misfolding is substantially increased relative to that in free solution, implying that this could be a primary cause of amyloid-like diseases such as Alzheimer's and Parkinson's disease (92-96). It is with this background that it is surprising that some proteins fold into their native structures even in such confined and crowded environments. Some theoretical and experimental studies have been reported for the GroEL–GroES bacterium chaperonin which indicates that the chaperonin encapsulates the synthesized polypeptide in a spherical cage and increases the kinetics of folding - a kinetic pathway is therefore thought to underpin this physical situation (97, 98).

Experimentally, the effect of confinement on protein stability was first studied by Eggers and Valentine (34). They found that the melting temperature of α-lactalbumin increased by approximately 32 $^{\circ}$C when it was entrapped inside a silica matrix through sol-gel encapsulation. This enhanced stability of confined proteins is of industrial interest since many biochemical reactions are catalyzed by enzymes that are immobilized inside a pore (33, 99, 100). Theoretical studies by Zhou and Dill showed that athermal confinement stabilizes a protein against reversible unfolding (36). Furthermore, they showed that the stabilization is highest when the size of the cavity is comparable to the radius of gyration of the folded state of the protein. Evidently, many open chain conformations are eliminated by confining a protein, thus leading to an entropically-driven stabilization of the folded state. Rathore *et al(101)* conducted Monte Carlo simulations using the Go model and showed that the folding temperature of the proteins studied increased upon confinement.

In contrast to these previous theories that only focus on entropic effects, typical experiments involve cavities that interact energetically with the protein. This is an important area of research

and many experimental studies have examined the structural changes that occur when a protein is adsorbed on a hydrophilic or a hydrophobic surface (102-107). To understand these results, one clearly needs to distinguish between the relative roles of energetic and entropic factors in determining the structural stability of confined proteins. In previous work from our group, we showed that placing a protein near an energetically neutral flat surface served to stabilize its folded state, due to the same entropic argument of Zhou and Dill (108). Surface hydrophobicity (λ) plays a non-monotonic role in this behavior with small hydrophobicities yielding increased stability and destabilization occurring at even larger hydrophobicities.

In the current paper, we study precisely these effects but in the context of a spherical pore through the device of Hydrophobic Polar lattice model (HP). We show that a neutral (or athermal) confinement stabilizes the protein against reversible unfolding, as suggested by Zhou and Dill. However, the introduction of surface hydrophobicity always serves to destabilize the folded state, so that energetic interactions always increase the propensity for unfolding. It is important to note that there is a range of surface hydrophobicity over which a spherical confinement induces stabilization relative to that in bulk solution, with this region being one where entropic stabilization dominates over the energetically induced destabilization. This result serves to qualitatively rationalize many previous experimental results, but most importantly this allows us a basis from which to understand our findings that the activity of alcohol dehydrogenase (ADH) increases when it is confined inside hydrophilic mesoporous silica (SBA-15) pores, while confinement in weakly hydrophobic methacrylate pores showed no essential reduction in its activity. These results are also in qualitative agreement with previous experimental findings by Sang *et al (109)* who reported increased enzymatic activity of lysosyme

and myoglobin when adsorbed inside hydrophilic SBA-15 pores and a drastic decrease in the activity when adsorbed inside propylated SBA-15 (hydrophobic) pores compared to that in free solution.

3.2 Materials and methods

3.2.1 Model

We modeled proteins in the frame work of the Hydrophobic-Polar lattice model introduced by Dill (110). In this "HP" model, individual amino acid residues are represented by beads labeled either as Hydrophobic (H) or Polar (P). The protein chain is grown as a self-avoiding walk on a three-dimensional cubic lattice. The bond length between two amino acid residues is held constant, equal to the length of the lattice that is used to discretize space. Water molecules are modeled in an implicit manner, and effective interactions are used to model the interactions between pairs of H-H, H-P and P-P residues. Two 'H' residues interact with an energy ϵ_{H-H} (ε_{H-H} = -1) =ε favoring formation of hydrophobic contacts when they are the nearest nonbonded neighbors. All other interactions ϵ_{H-P}, and ϵ_{P-P} are set to zero. The basis for setting the parameters as above is described below.

Since we deal with an incompressible mixture of H, P and solvent (s) molecules, the total energy of the system can be defined through the use of three independent energy interchange parameters, χ_1, χ_2 and χ_3. $\chi_1 = \epsilon_{\mathrm{HP}} - \frac{1}{2}[\epsilon_{\mathrm{HH}} + \epsilon_{\mathrm{PP}}]$ is the energetic cost involved in the formation of a H-P bond by breaking H-H and P-P bonds. In a similar way we define $\chi_2 = \epsilon_{\mathrm{Hs}} - \frac{1}{2}[\epsilon_{\mathrm{HH}} + \epsilon_{\mathrm{ss}}]$ and $\chi_3 = \epsilon_{\mathrm{Ps}} - \frac{1}{2}[\epsilon_{\mathrm{PP}} + \epsilon_{\mathrm{ss}}]$, as involving the interactions between an H group and

the solvent and the P groups and the solvent. Our choice of energy parameters implies that $\chi_1 = \chi_2 = \frac{1}{2}$, and $\chi_3 = 0$, implying that H-P and H-s interactions are equally unfavorable, with P-s interactions being athermal. It is important to emphasize that while this choice is somewhat arbitrary it emphasizes the fact that H groups only want to energetically be surrounded by other H groups. Further, this model is isomorphic to any choice of the different energy scales which yield the same $\chi_i's$. The introduction of the surface (S) sites yields three more energetic interchange parameters, but the point emphasized here is that our results are generically true across many situations where the hydrophobic monomers want to surround themselves with H monomers, while the S preferentially attracts the H monomers.

It is important to emphasize that, while the HP model is very popular and computationally expedient, it ignores the atomistic details of the molecules and also the specifics of water packing and the accompanying hydrophobic effects which are thought to have a significant influence on the free energy landscape of proteins, especially in their temperature dependence. This model also ignores Coulombic interactions, salt bridges and disulfide bonds. Despite these shortcomings, the HP model has proved to be an indispensable tool for understanding the physics of protein folding over the past two decades. The HP model mimics the behavior of real proteins with its rugged free energy landscape, unique thermal transitions and low degeneracy of the folded (native) state (53, 111). Studies show that the HP model protein folding is NP complete, indicating that the HP proteins have the same conformational search problem as real proteins (51, 112). In the current investigation, we study a 64- mer HP model protein designed by Yue *et al.(111)* The 64-mer protein has a well defined folded state forming a four-helix bundle with an

internal energy of 56ε. **Figure 3.1** presents two representative snapshots of the protein in its native state and unfolded state in a spherical pore of radius 4.

3.2.2 Simulation Details

Bulk simulations were carried out in a cubic box with periodic boundary conditions in all three directions. The lateral dimensions of the cell (x=65, y=65, z=65) are greater than the contour length of the chain studied, thus preventing any interaction of the chain with its periodic images. For the study of protein adsorption on flat surfaces, two impenetrable surfaces were modeled in the xy plane, one placed at z =0 and the other at $z = z_{max} = 65$, with periodic boundary conditions imposed in the x and y directions. The surface at $z = z_{max}$ is athermal, while the surface at $z = 0$ is modeled as a hydrophobic surface. Only the H groups of the protein interact with the hydrophobic surface when they reside on the $z = 1$ plane with an interaction energy –**λ**, where **λ** represents the hydrophobicity of the surface (**λ**=0 represents an athermal surface and **λ**=1 represents a surface attraction equal to the internal H-H interaction). The surface hydrophobicity can be tuned by varying **λ.**

Simulations were also carried out for confined proteins placed inside an approximate spherical cavity in a cubic lattice. To define this cavity, we first drew a sphere in free space, placed it on the lattice and then only included all those sites that are either inside the sphere or those sites from which bonds that emanate are "cut" by the sphere. By our convention, all the H units which are within a distance of √3 from the surface interact with the surface lattice sites with interaction energy -**λ**. So, on a flat surface, any given H residue can interact with 9 surface lattice sites. As

the curvature increases, the number of surface sites that a given residue can interact with increases as shown in **Figure 3.2**.

Figure 3.3 shows the average number of surface sites interacting with a given residue at a distance of √3 unit from the surface plotted against 1/R (R is the radius of the cavity). It is evident that, as the size of the cavity decreases, the average number of surface lattice sites available for interaction increases. The number of interaction sites decreases linearly with (1/R) and approaches the well-documented flat surface behavior for a cavity of infinite radius. The important emergent point, which shall play a critical role in the role of surface energetics in protein behavior in cavities is that, surface energetics progressively becomes a stronger effect for smaller cavities.

Multiple canonical Monte Carlo simulations (61, 68) were carried out at different temperatures T*($=k_BT/\epsilon$) and surface hydrophobicities (**λ**). We have employed local moves i.e., internal flip, end bond flip and crankshaft flip along with the more non-local configurational bias Monte Carlo moves (113) to equilibrate the proteins. We used non-local moves since simulations carried out using only local moves got trapped in metastable states and thorough sampling of the phase space was not possible. The simulation data was combined using the Weighted Histogram Analysis Method (WHAM) (114) method as described by Kumar et al (115) to obtain the density of states (DOS), and from there the free energy of the system (see below). Thermodynamic averages of the specific heat and internal energy were then calculated using the DOS.

We considered protein folding as a two state process. A protein conformation with the ground state energy of 56ε is considered to be in the folded state (also called the native state) and all

other states are considered to be unfolded. The temperature at which the free energy of the folded state is equal to that of the unfolded state is called the folding temperature or melting temperature of the protein. Another estimate of the folding temperature is obtained from the temperature dependent specific heat. The specific heat estimate, which is based on energy fluctuations, does not explicitly define folded or unfolded configurations, and is therefore less model dependent than the free energy estimate. The temperature at which there is a maximum in the specific heat curve corresponds to the folding temperature of the protein. The specific heat based estimate of the folding temperature is always higher than the free energy estimate possibly due to finite size effects, or due to our specific free energy definition of the folding transition temperature. Note however that, both the estimates agree qualitatively so that either estimate is reliable in this context.

3.2.3 Weighted Histogram Analysis Method (WHAM)

The thermodynamic average of an observable, A, which is a function of U (U is the intramolecular energy of the protein in the present study) at any given temperature can be calculated mathematically using Eq 1, where A(U) is defined as the average of A over all configurations with an energy U and Ω (U) - the density of states (DOS) - represents the number of different configurations which have an energy U.

$$\langle A \rangle_T = \frac{\int dU \Omega(U) \exp(-U / k_B T) A(U)}{\int dU \Omega(U) \exp(-U / k_B T)} \qquad \textbf{(1)}$$

Ω(U) is independent of temperature and is sufficient to calculate the thermodynamic averages of the system at any given temperature. This requires calculation of Ω(U) for all possible values of

U through thorough sampling of phase space. Since the free energy landscape of proteins is very rugged due to the existence of many metastable states, canonical Monte Carlo simulations were carried out at various temperatures (T^*) and surface hydrophobicity **(λ)** so as to thoroughly explore phase space. These simulations give us the local free energy landscape of the system. All the data from these simulations were combined using the WHAM to get the global free energy landscape. The DOS and free energy of the system are estimated directly from WHAM using a self consistent iterative process represented by Eqs 2 and 3.

$$\Omega_m = \frac{\sum_{k=1}^{K} g_k^{-1} H_{mk}}{\sum_{k=1}^{K} g_k^{-1} N_k \Delta U \exp\left[f_k - \beta_k U_m \right]} \tag{2}$$

$$f_k = -\ln \sum_{m=1}^{M} \Omega_m \Delta U \exp(-\beta_k U_m) \tag{3}$$

Here, K is the total number of simulations, $\boldsymbol{\Omega_m}$ is the DOS of the state *m*, $\boldsymbol{f_k}$ is the free energy of the k^{th} simulation, and $\boldsymbol{g_k}$ is the statistical uncertainty associated with k^{th} simulation. $\mathbf{g_k = 1 + 2*\tau}$, where $\boldsymbol{\tau}$ is the area under energy autocorrelation curve**.** $\boldsymbol{H_{mk}}$ is the histogram entry of the m^{th} state from the k^{th} simulation. $\mathbf{N_k}$ is the total number of histogram entries from the k^{th} simulation. $\boldsymbol{\beta_k}$ is the inverse of the thermodynamic temperature ($= \frac{1}{k_B T}$) of the k^{th} simulation. U_m is the energy associated with the m^{th} state and **Δ**U is the width of the histogram bins, which is 1 in our current study. Eqs. 2 and 3 are solved iteratively for self-consistency to estimate the DOS and free energy of the system.

It should be noted that the exclusion of g_k or incorrect estimates of g_k result in an inaccurate estimation of the DOS and free energy. The effect of g_k manifests itself prominently for simulations carried out near the folding temperature of the protein. At these temperatures, the system takes a very long time to relax and hence simulations should be carried out considerably longer than the appropriate system relaxation times to get the correct estimate of g_k. We point here to our previous work, where we did not perform long enough simulations at low temperatures (2×10^8 MC cycles). Instead, the value of g_k was simply estimated based on higher temperature runs. In the current paper we have carried out simulations for 8×10^8 MC cycles, consisting of 4×10^8 equilibration cycles and 4×10^8 production cycles. While the revised, more reliable analysis, does not qualitatively alter the conclusions drawn on the effect of hydrophobic flat surfaces on protein stability, relatively large quantitative effects persist. In summary, the new results (which are now accurate to the reliability of our extensive simulations) always show a relative small stabilization of the folded state, as will be discussed below. The instantaneous normalized energy auto correlation function is given by

$$C(t) = \sum_{\tau=1}^{N} \frac{\left(U(t) - \langle U \rangle\right) - \left(U(t+\tau) - \langle U \rangle\right)}{\left(U(t) - \langle U \rangle\right)\left(U(t) - \langle U \rangle\right)}$$

Near the folding temperature of the protein, it is very hard to overcome the free energy barriers since the Boltzmann acceptance probability at these temperatures is small. As a consequence, simulations carried out for short times underestimate the value of C(t), which would result in large errors in the estimation of the DOS of the system.

3.2.4 Experimental details

All materials and reagents were used as received. Alcohol dehydrogenase (ADH), glutaraldehyde (GA), **β**-nicotinamide adenine dinucleotide (NADH), isobutaldehyde, thiamine pyrophosphate (ThDP), and buffer salts were purchased from Sigma-Aldrich Chemicals (Milwaukee, WI). SBA-15 was synthesized according to published protocols. Methacrylate resin was purchased from Itochu Chemicals America (Relizyme HA403/S, White Plains, NY).

96 well plate assay conditions: 105 µL buffer A (50 mM potassium phosphate buffer, pH 6.8, 2.5 mM MgSO4, 0.1 mM ThDP), 20 µL NADH 2.5 mM in buffer A (final concentration 0.25 mM) and 25 µL 25 mg particles/mL buffer A were used in the experiments. The assay was started by adding 50 µL isobutaldehyde 120 mM in buffer A (final concentration 30 mM) to each well. ADH converts isobutaldehyde to isobutanol using NADH as a cofactor; NADH absorbs light at 340 nm and NAD^+ does not. The kinetic activity was followed by observing the decrease in absorbance (i.e. NADH) over time. The conditions in each well were kept constant in order to obtain statics of technical replicates (n=8)**.**

SBA-15 immobilization: SBA-15 particles (25 mg) were mixed on an end-over-end mixer with 1 mL (1 mg ADH/mL buffer A) solution for 4 hours. Subseuqently, the particles were centrifuged out and the supernatant assayed at 280 nm on a nanodrop UV-Vis spectrophotometer (Thermo Fisher Scientific, Pittsburgh, PA) to determine the solution protein concentration. SBA-15 particles were washed with fresh buffer A, and then assayed according to the 96 well plate assay conditions described above. The kinetics of the immobilized enzyme were then compared to

those of equivalent enzyme free in solution so as to delineate the role of surfaces on enzyme activity.

Methacrylate resin immobilization: Methacrylate resin (50 mg) was rehydrated in an end-over-end mixer with deionized H_2O (diH2O) for 1 hour, centrifuged and resuspended in 1 mL (0.1% GA) and mixed similarly for 1 hour. The resin was washed five times with diH2O to remove excess GA. The resin was then mixed with 1 mL (10, 5, 2.5, 1, or 0 mg ADH/mL buffer A) solution for 2 hours. Resin particles were centrifuged and the supernatant assayed at 280 nm to determine protein concentration. Methacrylate resin was washed with fresh buffer A, then assayed according to the 96 well plate assay conditions described above. The kinetics of immobilized enzyme was then compared with those of equivalent enzyme free in solution.

3.3 Results and Discussion

3.3.1 Simulation results

3.3.2 a) Athermal Surfaces: The stability of the 64-mer HP protein in the presence of a flat surface and inside a cavity was compared with its stability in the bulk. Near a flat, energetically neutral surface (λ=0) the protein behaves as in the bulk, apparently because it has no propensity to adsorb to the surface. We now consider spherical cavities of progressively smaller volumes. **Figure 3.4** shows the specific heat curve of the 64-mer HP protein at a surface hydrophobicity λ=0 for different geometries. The folding temperature of the protein is estimated by the peak in the specific heat curves. It is evident from **Figure 3.4** that, as the degree of confinement increases, the peak in the specific heat curve shifts to higher temperatures indicating higher thermal stability of the protein under confinement. This result is in agreement with many past

theoretical works (37, 99, 116, 117) and can be explained by the fact that, upon confinement, the unfolded state of the protein loses considerably more entropy compared with that of the compact folded state. (**Figure 3.5**)

3.3.3 b) Hydrophobic Surfaces: **Figure 3.6** shows the difference in free energy of the folded state and unfolded state of the protein at different values of surface hydrophobicity (**λ**) calculated at the melting temperature of the protein near a neutral confining surface. In the case of a flat surface, the free energy decreases with an increase in **λ** for $\lambda < \lambda_C$ (λ_c=0.32). Since the protein adsorbs onto the surface in its folded state for **λ**>0.15, for these $\lambda < \lambda_C$ values the surface energetics apparently increase its stability. These effects are qualitatively consistent with our previous work, but the current results are quantitatively much smaller than the previous one by about $0.7k_BT$ for the reasons explained earlier (see WHAM section). For $\lambda > \lambda_C$, $\Delta G > 0$; indicating that the folded state is destabilized relative to the folded state near a neutral surface due to energetic effects. We presume that surface interactions between the H groups of the protein and the surface overcome the internal energy of the protein. The protein then unfolds under these conditions. Although the protein loses many internal H-H contacts in this process (which increases its internal energy), there is an overall decrease in free energy because of surface interactions.

For the case of a cavity, we do not observe an additional stabilization of the folded state on the inclusion of surface energetics. Since the protein is dramatically stabilized by a spherical cavity for λ=0, we can still observe that the protein is stabilized relative to the free solution especially for weakly hydrophobic surfaces (**λ** < 0.23) reflecting the dominance of entropic effects in these cases. For large enough values of **λ**, confinement always destabilizes the protein relative to the

bulk due to protein-surface interactions (**Figure 3.7**). These are the central conclusions of our work and they arise due to interplay between the entropic and enthalpic dominance at small and large hydrophobicities respectively

3.3.4 Experimental results

The kinetic activity of ADH immobilized non-covalently inside SBA-15 resin pores (hydrophilic) and methacrylate resin pores (weakly hydrophobic) was compared to its activity in free solution. We found that the kinetic activity (indicated by the initial slope of the absorbance curve) was drastically increased when the enzyme was immobilized inside hydrophilic SBA-15 pores (**Figure 3.8**) whereas the activity was relatively unaffected at different enzyme concentrations inside methacrylate resin pores (**Figure 3.9** and **Table 3.1**). These results lend considerable weight to our theoretical findings. In the case of hydrophilic pores, the protein has no energetic propensity to unfold. Additionally, the entropic factors suggested by Zhou and Dill lend additional stability to the folded state. In combination, these effects rationalize our findings for SBA-15. The results of ADH inside methacrylate pores (water contact angle for PMMA is 73°) are also consistent with our findings (**Figure 3.9**). For these cases, entropic stabilization is nearly compensated by surface energetics leading to a protein whose stability is unaffected compared with the bulk. Thus while we cannot estimate the value of $\boldsymbol{\lambda}$ for these pores, and especially because the simulations consider spherical cavities while the experiments relate to cylindrical pores, we conjecture that it must be close to the value where the entropic and energetic factor balance each other.

3.4 Discussion

The relationship between the structure of the enzyme and its activity is still not well established, since enzymatic activity depends on the accessibility of the active site of the enzyme by the substrate. Therefore changes in the native structure of an enzyme may lead to either an increase or decrease in its activity depending on the location of the active site in the enzyme as detailed in the article by Tran et al (50). Here we want to point out to the work by Sang et al, who observed an increased enzymatic activity of lysozyme and myoglobin when adsorbed inside hydrophilic SBA-15 pores while the activity drastically decreased when adsorbed inside propylated SBA-15 (hydrophobic) pores relative to that in free solution. Furthermore, analysis of the structure of the adsorbed enzymes showed that, while lysosyme gained secondary structure (β-sheet content) inside hydrophilic pores, myoglobin lost some percentage of its α-helix content inside hydrophilic pores. In contrast, there was a drastic disruption of the secondary structure upon adsorption inside hydrophobic SBA-15 pores for both enzymes. Although, our experimental findings are in agreement with the simulations results, it would be premature to draw general conclusions due to the following caveats. The HP model for protein is very simplistic and neglects atomistic details and interactions. The 64-mer HP protein forms four-helix bundles in its native state, whereas ADH is known to be made of helices, sheets and random coils. The effect of confinement and surface energetics are less pronounced inside cylindrical SBA-15 pores than in spherical pores and hence a direct relationship between the studied enzyme and our model is difficult to establish. With all of these caveats, nevertheless, our study explains the general physics of protein folding and unfolding inside athermal and hydrophobic pores.

Table 3.1: Enzymatic activity of ADH immobilized at different surface concentrations inside a methacrylate resin pores

Protein Loading	Enzymatic activity *
3µg / mg resin	-3.1 ± 1.8
7µg / mg resin	-2.4 ± 0.7
15µg / mg resin	-2.6 ± 0.5
17µg / mg resin	-3.3 ± 0.4

*K_{cat} x10^{11} (1 / sec)

3.5 Conclusion

In the current paper, we have shown that a neutral spherical confinement stabilizes the protein against reversible unfolding. The unfolded state loses more entropy upon confinement compared to the folded state resulting in higher stability of the protein under neutral confinement. Inclusion of energetic effects, always serves to destabilize the protein relative to this athermal case. However surface hydrophobicity has to increase to a relatively strong value where the energetics can overcome these entropic stabilizing effects. The protein is destabilized relative to the bulk solution for any larger λ. Our results allow us to rationalize the increased activity of ADH inside the hydrophilic cavities of SBA-15, where entropic effects probably dominate. Perhaps more interesting are our results on the weakly hydrophobic pores of a methacrylate resin, where the hydrophobic energetic destabilizing factors, appears to almost balance the entropic stabilizing forces leading to essentially no change in protein activity upon confinement. It appears that the proper tailoring of energetic effects in spherical confinement can be employed to change the activity of a protein at will.

3.6 Stability of hydrophilic surfaces

Figure 3.10 shows the effect of surface hydrophilicity on the folding temperature of the protein under different geometries. It is evident that for a neutral surface (λ=0), the protein inside a spherical cavity (R=4 and R=6) has a higher folding temperature indicative of higher thermal stability than compared to the on a flat surface and on a positive curvature surface. This is purely an entropic effect as discussed earlier. Confining proteins helps in eliminating many open chain conformations of the unfolded state of the protein due to physical constraints. Consequently this loss in entropy of the unfolded state drives the equilibrium towards the folded state. This is further reassured from the fact that the folding temperature inside cavity of radius 4 is higher than inside a cavity of radius 6. The protein on a flat surface and on the outside of the pore behave similar to as in bulk since at a value of λ=0, they do not have any propensity to adsorb on to the surface and further due to a large volume available outside they do not feel any confinement due to the presence of the surface. The folding temperature of the protein for an athermal surface with different geometries is listed in **Table 3.2**.

Table 3.2: Folding temperature of the 64mer protein under different degree of athermal confinement

Geometry	Folding temperature (T*)
Radius 4 (negative curvature)	0.30
Radius 6 (negative curvature)	0.295
Flat surface	0.29
Radius 10 (positive curvature)	0.29

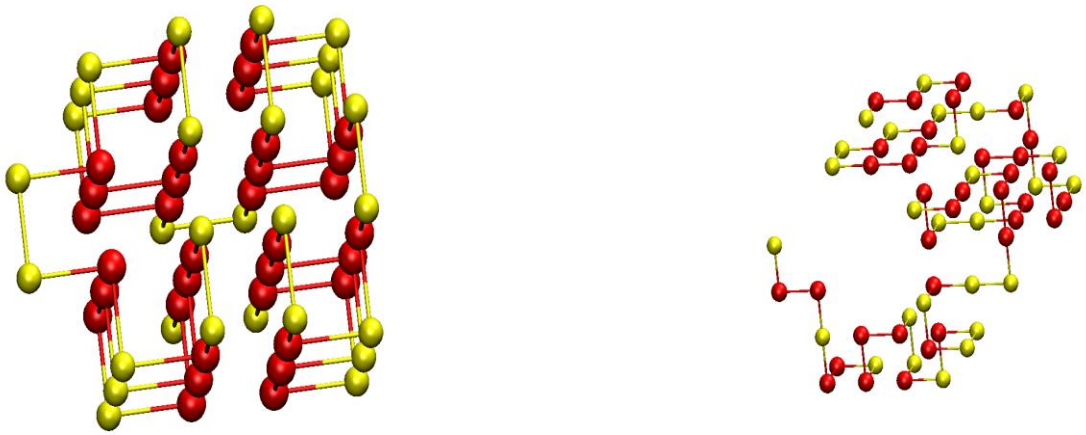
Now as the value of surface hydrophilicity (λ) is gradually increased, we see that it has a substantial effect on the folding temperature of the protein depending on the geometry which can be explained as follows:

a) Adsorption on the inside (negative curvature): We see contracting behavior of the protein depending on the size of the cavity. The protein inside the cavity of radius 4 is dramatically destabilized by the increase in surface hydrophilicity (λ). This is explained by the fact that, inside smaller cavities surface energy dominates entropy. As a result, the protein denatures (unfolds) to gain protein-surface contacts indicated by a decrease in the folding temperature. On the other hand we observe that the folding temperature of the protein increases with the increase in the value of λ upto a critical value of λ=0.3 from T_m=0.295 to T_m=0.32 and then decreases. For the values of λ > 0.52, the protein is destabilized even relative to its bulk behavior. This behavior can be explained as follows. For low values of λ, the protein gains stability due to two independent factors. One from the entropic stability from the confinement and second from the energetic stability due to interactions between the 'P' (polar) groups of the protein and the surface. The surface energetic are strong enough to stabilize the folded state but not strong enough to denature the protein, ie the energetic gain by unfolding is countered by the energetic gain upon adsorption in the folded state and the entropic stabilization. For values of λ > 0.52 the surface energetics become strong enough to rupture the protein and hence the protein becomes destabilized.
b) Adsorption on flat surface: The folding temperature of the protein increases with increase in value of surface hydrophilicity from T=0.29 for λ=0 to T=0.31 for λ=0.64 and

then decreases with the further increase in λ. The percentage increase in the folding temperature (around 6%) is less compared to that of the adsorption on the inside of a cavity of radius 6 (around 9%) due to absence of entropic stabilization of confinement.

c) Adsorption on the outside of the surface: Maximum percentage increase in the folding temperature (around 13%) is observed in the case of adsorption of the protein on the outside of a hydrophilic surface with positive curvature. This increased stabilization comes from the two dependent factors. Due to weaker protein surface interactions the protein can be stabilized at higher values of λ compared to that on a flat surface and on the inside of a cavity (negative curvature). Further, due to more volume available for the protein near the surface due to its positive curvature the protein in its folded state can expose relatively more surface area (more 'P-S (surface) contacts) resulting in higher stabilization. For very high values of λ though, the protein denatures due to strong protein surface interactions.

Figure 3.11 shows the free energy difference between the folded and unfolded state $(\Delta G_{folded} - \Delta G_{unfolded})$ of the protein compared to its bulk free energy under different geometries as a function of surface hydrophilicity (λ) at their respective melting temperatures. A negative value of the free energy difference indicates the stabilization of the folded state of the protein relative to its stability in the bulk. **Figure 3.12** confirms the findings of **Figure 3.11**. In the case of radius 4 (inside), the free energy difference is always positive since the unfolded state becomes the minimum free energy state whereas in all other cases the folded state is stabilized upto a critical value of λ depending on the geometry studied.



A) *B*)

Figure 3.1: Snapshots of a 64-mer HP protein obtained through canonical Monte Carlo simulations in a spherical cavity of radius 4. Yellow spheres represent polar (P) residues and red spheres represent hydrophobic (H) residues. (A) Native state of the protein - forming four-helix bundles with the ground state energy of 56ε. B) Unfolded state of the protein with an energy of 25ε.

A) *B*)

Figure 3.2: The effect of surface curvature on the number of surface sites available for interaction with a given hydrophobic (H) residue of the protein which is at a distance of √3 from the surface sites. Red sphere represents a hydrophobic residue of the protein and silver spheres represent surface sites. A) On a flat surface, the 'H' residue interacts with 9 surface sites while B) Inside a cavity of radius 4, the 'H' residue mostly interacts with 12 surface sites.

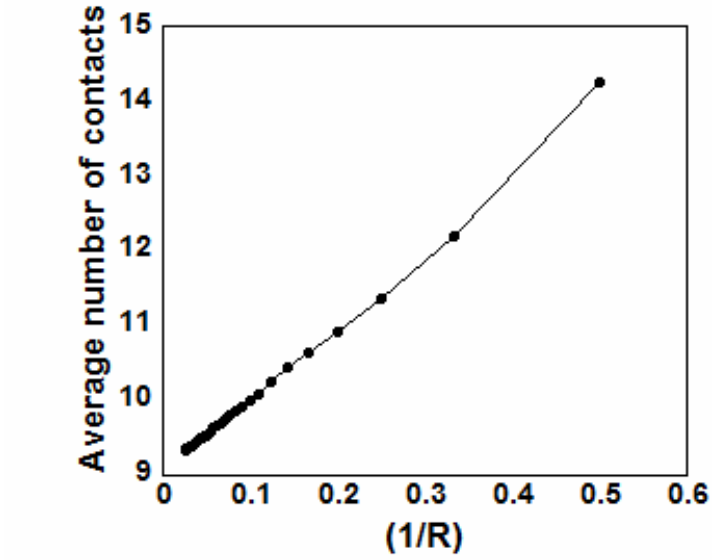


Figure 3.3: Average number of surface sites with which any 'H' residue of the protein can interact when it is at a distance of √3 from the surface as a function of the inverse of the cavity radius. (On a flat surface, the 'H' residue can interact with 9 surface sites)

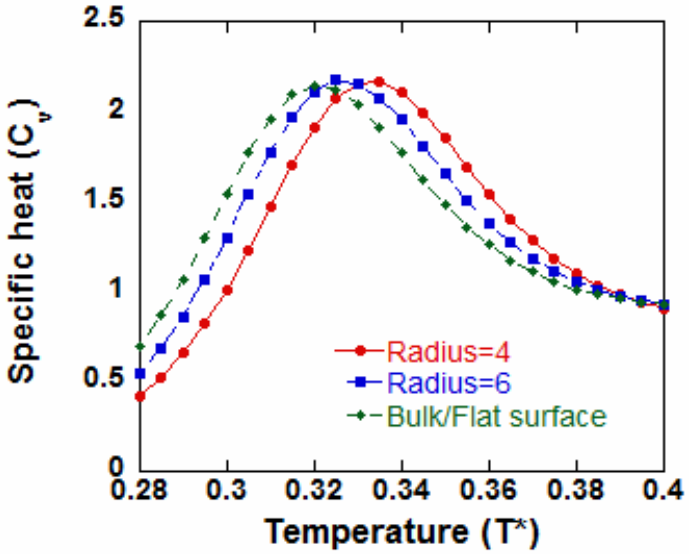


Figure 3.4: Specific heat curve of the 64-mer HP protein at different degrees of athermal confinement, both at a flat surface and inside spherical cavities of radius 4 and 6 respectively

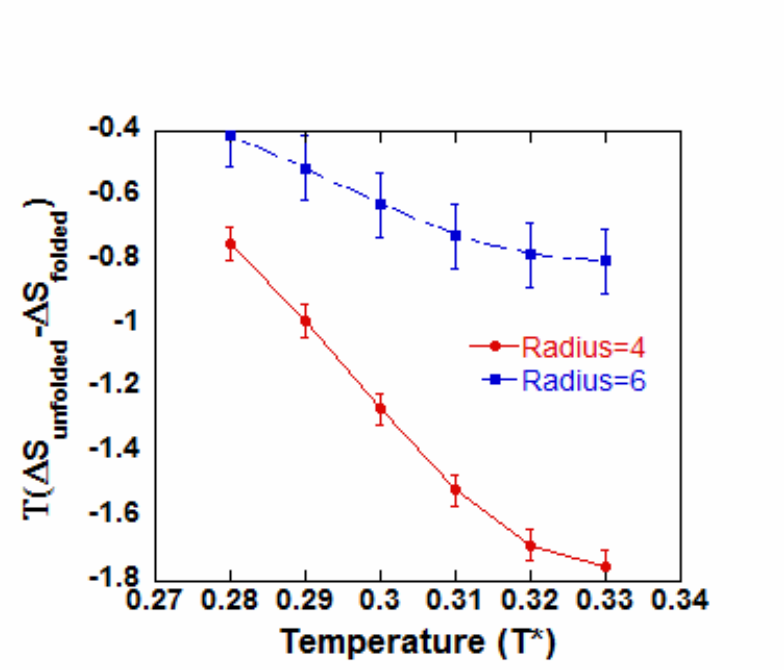


Figure 3.5: Entropic difference between the unfolded and the folded state of the 64-mer HP protein inside energetically neutral (λ=0) spherical pores as a function of temperature

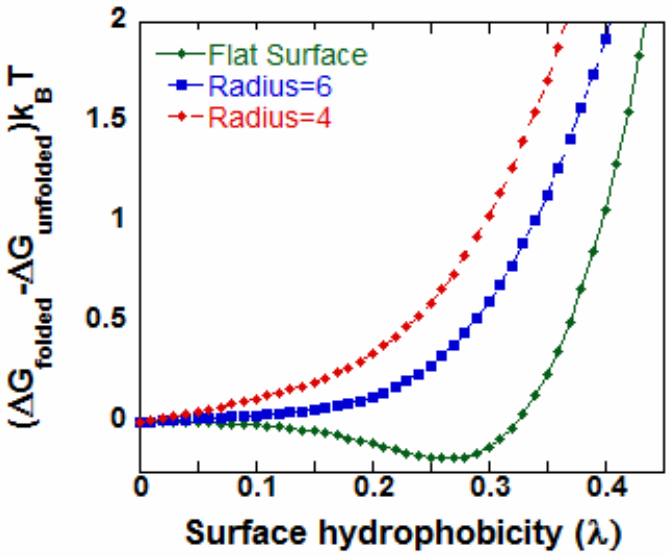


Figure 3.6: Free energy difference between the folded and unfolded state of the 64-mer HP protein with varying surface hydrophobicity (λ) at different degrees of confinement calculated at the melting temperature of the protein at λ=0, of the confining surface. (T*=0.29 for flat surface, T*=0.295 for a cavity of radius 6, T*=0.302 for a cavity of radius 4).

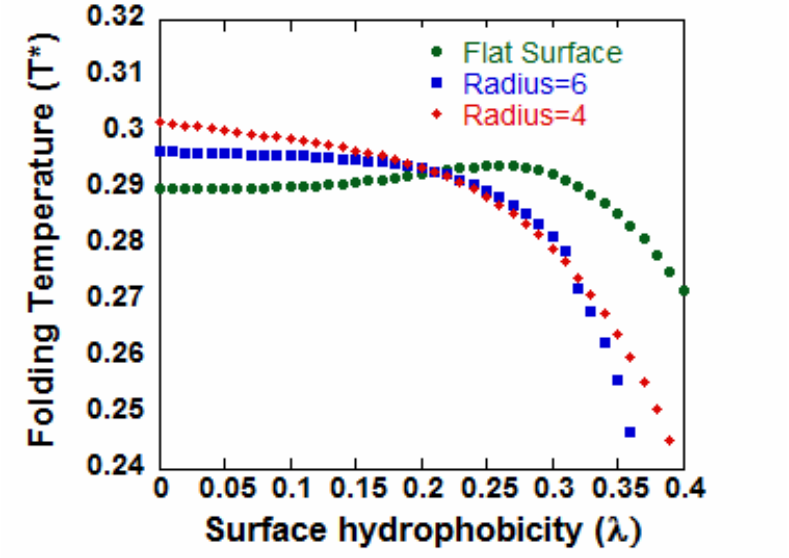


Figure 3.7: Folding (melting) temperature (T_m) of the 64-mer HP protein as a function of surface hydrophobicity (λ) for different degrees of confinement

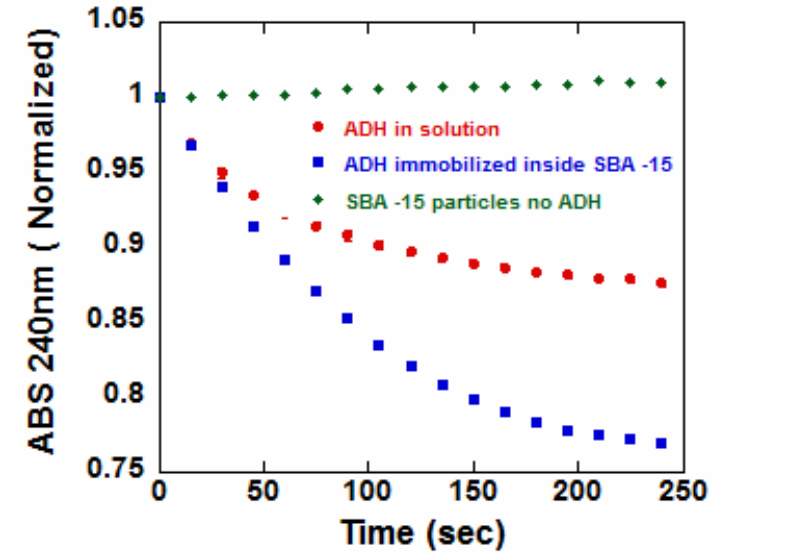


Figure 3 8: Enzymatic activity of alcohol dehydrogenase (ADH) non-covalently immobilized on mesoporous silica (SBA-15).

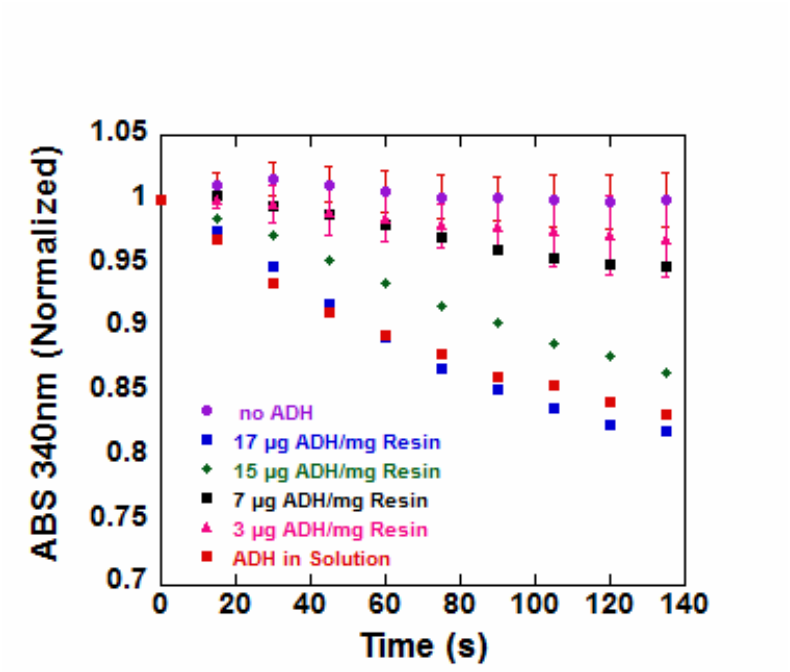


Figure 3.9: Enzymatic activity of alcohol dehydrogenase (ADH) non-covalently immobilized on methacrylate resin**.**

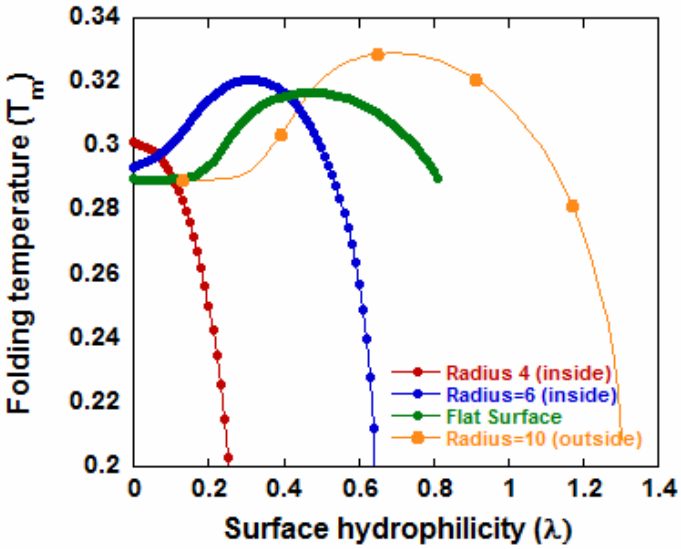


Figure 3.10: Folding (melting) temperature (T_m) of the 64-mer HP protein as a function of surface hydrophilicity (λ) for different degrees of confinement

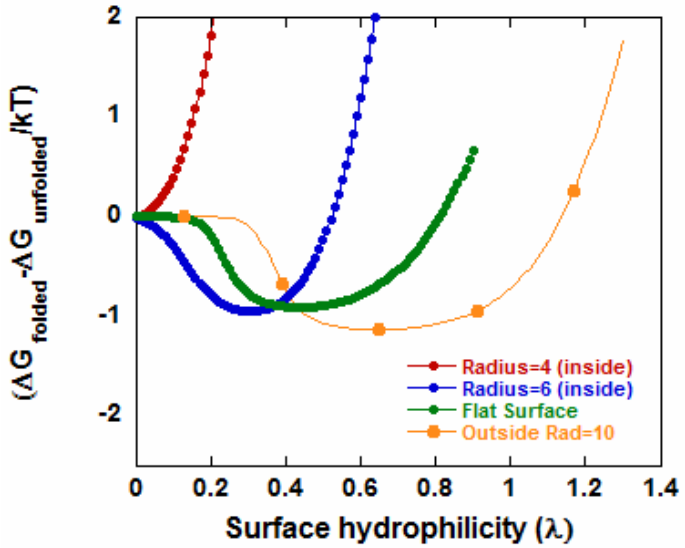


Figure 3.11: Free energy difference between the folded and unfolded state of the 64-mer HP protein with varying surface hydrophobicity (λ) at different degrees of confinement calculated at the melting temperature of the protein at λ=0, of the confining surface. (T*=0.29 for flat surface, T*=0.295 for a cavity of radius 6, T*=0.302 for a cavity of radius 4,T*=0.29 for adsorption outside of the cavity)

Chapter 4

Surface-Mediated Protein Disaggregation

Preventing protein aggregation is of both biological and industrial importance. Inter-protein interactions between the hydrophobic residues of the protein are known to be the major driving force for protein aggregation. In this chapter we show how surface chemistry and curvature can be tuned to mitigate these inter-protein interactions. Our results calculated in the framework of the Hydrophobic–Polar (HP) lattice model show that inter-protein interactions can be drastically reduced by increasing the surface hydrophobicity to a critical value, corresponding to the adsorption transition of the protein. At this value of surface hydrophobicity, proteins lose inter-protein contacts to gain surface contacts, and thus the surface helps to reduce the inter-protein interactions. Further, we show that the adsorption of the proteins inside hydrophobic pores of optimal sizes are most efficient both in reducing inter-protein contacts and simultaneously retaining most of the native-contacts probably due to confinement-induced stabilization.

4.1 Motivation

Enzyme immobilization is of great importance in industrial applications like bio-fuel cells, biosensing and recognition, drug delivery and bio-catalytic films because this makes for a repeated usability of these catalysts (50, 118-121). However, a major concern with surface immobilization is that the enzyme loses its activity. This loss in activity is attributed to a change in the structure of the enzyme from the native, active form to an unfolded, inactive structure. The two major factors driving these conformational changes of proteins on surfaces are protein-surface interactions and inter-protein interactions. While these facts appear to be universally accepted, an interesting new result emerged from the experimental work by Asuri et al (122)

which showed that enzymes immobilized on single walled carbon nanotubes (SWNTs) retained more activity compared to ones adsorbed on flat supports under denaturing conditions. They rationalized these findings by attributing the increased stability on SWNTs to reduced inter-protein interactions since the center to center distance between the immobilized enzymes increases on SWNTs compared to that on flat supports at the same surface adsorption density. Other experiments have shown that proteins adsorbed onto surfaces with patchy domains retained more activity than on homogeneous domains, again due to reduced inter-protein interactions (123). This concept of preventing protein aggregation is also of biological importance in preventing autoimmune diseases such as Alzheimer's and Parkinson's disease (124, 125).

In a related point we note that many experiments emphasize the role of chaperonins in protein repair. One of the major steps in the process of repair is to disaggregate aggregated proteins and then create a shielded environment for the resulting individual proteins to refold into their native state (126, 127).

Driven by these apparently disparate ideas, in the current paper we address the following questions: (a) How does surface curvature (positive and negative) affect inter-protein interactions? (b) How can we tune surface behavior to facilitate protein disaggregation? Our simulations carried out in the framework of the Hydrophobic Polar (H-P) lattice model show that inter-protein interactions can be dramatically reduced by increasing the surface hydrophobicity for all adsorbing geometries, i.e., surfaces with positive or negative curvature. In our calculations we treat two proteins and ask which state is more favorable – two aggregated proteins or two disaggregated proteins. For this particular equilibrium we find that the optimal curvature, which

is negative (corresponding to a protein confined inside a pore), at which protein disaggregation is optimized. This optimum is driven by a competition between two effects: as we reduce the radius of a pore, the chains more readily have protein-surface interactions. Thus, they are able to reduce inter-protein interactions leading to protein disaggregation. On the other hand, with decreasing radius the law of mass action (or an increase in protein concentration) drives the chains to aggregate. These findings also shed light on the role of chaperonins in protein disaggregation as discussed above (87, 88).

4.2 Simulation Methods

4.2.1 Model

We used the Hydrophobic Polar (H-P) lattice model designed by Dill (110), which is motivated by the observation that hydrophobic interactions between the amino acid residues are the major driving force for proteins to fold into their native state (2) In this model individual amino acid residues are labeled either as Hydrophobic (H) or Polar (P) – the chain is grown as a self-avoiding walk in a three dimensional cubic lattice. No two residues can occupy the same lattice site. The bond length between the adjacent amino acid residues is held constant and is equal to the length of the lattice used to discretize space. Water molecules are modeled in an implicit manner. To account for this implicit solvent, an effective attraction is introduced between the non-bonded H groups, $-\epsilon_{H-H}$ (ϵ_{H-H}=1), when they are the nearest neighbors on the lattice (coordination number = 6). All binary interactions involving a polar group, i.e., ϵ_{H-P}, and ϵ_{P-P}, are set to zero. The basis for selecting these energy parameters is detailed in our previous work (128). The HP model mimics the behavior of real proteins with its rugged free energy landscape,

unique thermal transitions and low degeneracy of the folded (native) state. In the current study we have studied a 42mer HP protein designed by Yue et al. with ground state energy of -34 ε_{H-H} (53, 111). The protein used in the current study has 34 native contacts in its folded state. Native-contacts are a measure of the intra molecular 'H-H' contacts of the protein in its folded state which defines the secondary structure of the protein.

4.2.2 Simulation details

Bulk simulations were carried out in a cubic box with periodic boundary conditions imposed in all the directions. The lateral dimensions are chosen such that they are significantly larger than the radius of gyration of the unfolded state of the protein (≈5), thus minimizing any interactions between the periodic images. For the study of protein adsorption on flat surfaces, two impenetrable surfaces were modeled in the xy plane one placed at z=0 and the other at $z = z_{max}$ with periodic boundary conditions imposed in x and y directions. The surface at $z = z_{max}$ is athermal, while the surface at z=0 is modeled as a hydrophobic surface. Only the 'H' residues of the protein interact with the hydrophobic surface when they reside on the z=1 plane with an interaction energy $-\lambda$, where λ represents the hydrophobicity of the surface (λ =0 represents an athermal surface and λ =1 represents a surface attraction equal to the internal H-H interaction). Note that, in contrast to the bulk, the group at the surface interacts with all surface neighbors which are within a distance of √3, i.e., 9 neighbors. The surface hydrophobicity can be tuned by varying λ. The total energy of the system is the sum of three components namely the intramolecular energy, protein-surface interaction energy and inter-protein interaction energy.

Simulations on the inside (negative curvature) and on the outside (positive curvature) of a spherical pore were carried out by approximately forming a spherical cavity in a cubic lattice as described in our previous work. By our convention, all the 'H' residues which are within a distance of √3 from the surface interact with the surface lattice sites with interaction energy - λ. So, as the curvature increases, the number of surface sites that a given 'H' residue can interact

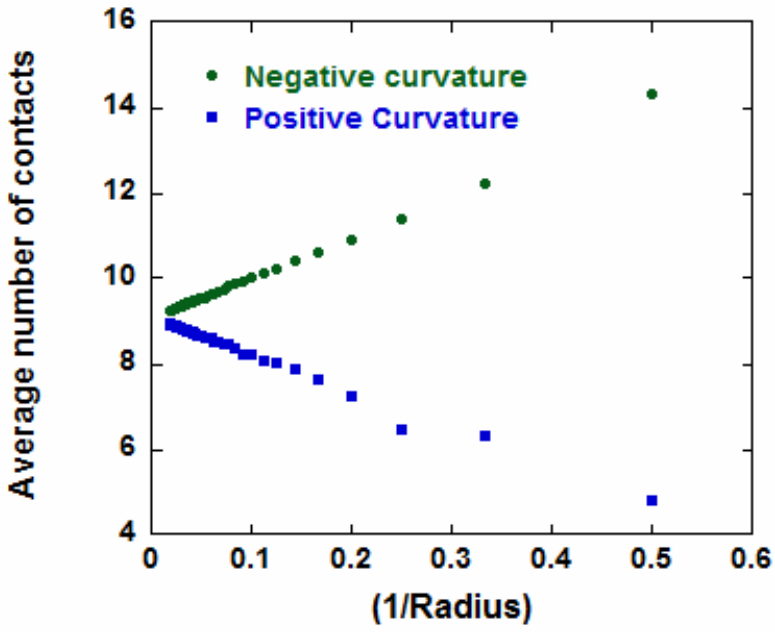


Figure 4.1: Average number of surface sites with which any 'H' residue of the protein can interact when it is at a distance of √3 from the surface as a function of the inverse of the cavity radius. (On a flat surface it can interact with 9 surface sites)

with increases/decreases depending on whether the surface curvature is negative/positive. **Figure 4.1** shows the average number of surface sites interacting with a given 'H' residue at a distance of √3 unit from the surface plotted against 1/R (R is the radius of the cavity). It is evident that the number of interaction sites decreases/increases linearly with (1/R) for surfaces with

negative/positive curvature and approaches the well-documented flat surface behavior for a cavity of infinite radius. The important emergent point, which shall play a critical role in the role of surface energetics on proteins is that surface energetics becomes progressively a stronger/weaker effect depending on the curvature of the surface – these effects become stronger for negative curvature (i.e., for proteins confined in a pore). Multiple canonical Monte Carlo simulations were carried out at different temperatures T*($= \frac{Tk_B}{\varepsilon}$) and surface hydrophobicities (λ) to sample the phase space thoroughly. Weighted Histogram Analysis Method (WHAM) was used to calculate the density of states of the system (DOS) from which thermodynamic averages like specific heat, total energy are calculated.

4.3 Results and Discussion

4.3.1 Protein aggregation in bulk

The effect of inter-protein interactions on the stability of proteins was studied by considering a system with two protein chains of the 42mer model protein. The model protein exhibits protein-like folding and a collapse transition at T*=0.26 and T*=0.49, respectively, as shown from the specific heat curves (**Figure 4.2**). These systems were studied under the following conditions.

(i) Isolated proteins (Dilute Regime): To mimic the behavior of single proteins, we consider two proteins, but turn off all inter-protein interactions

(ii) Non-isolated proteins (Concentrated regime): Inter-protein 'H-H' interactions is exactly equal to their intra-molecular values

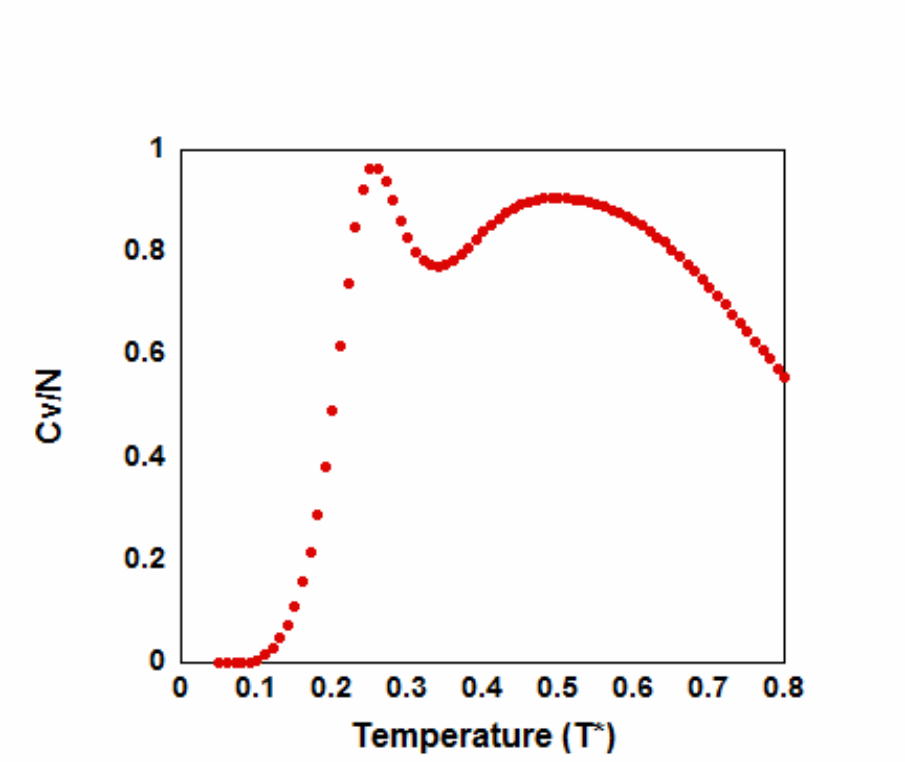


Figure 4.2: Specific heat curve of the 42mer HP protein as a function of temperature (T*). The peaks at T*=0.49 and T*=0.26 represent the collapse and folding transition of the protein respectively.

Figure 4.3a shows the total energy of the two 42mer protein chains in the bulk under isolated and non-isolated conditions. We observe that the total energy of the system with inter-protein interactions is always lower than the isolated case. **Figure 4.3b** shows the intra-molecular energy of the proteins as a function of temperature. It is evident that the intra-molecular energy of proteins in the non-isolated case is higher (less favorable) than in the isolated case, at least over the temperatures we studied. Apparently, inter-protein interactions prevent these proteins from folding into their native state by stabilizing the unfolded, but aggregated states. These studies imply that the aggregated state is always the minimum free energy state as also pointed out through previous biological experiments (129). Although the protein loses many intra-molecular

'H-H' contacts, it makes up for it through inter-molecular 'H-H' interactions. For the protein in study the percentage loss of native contacts at T*=0.36 is around 20%.

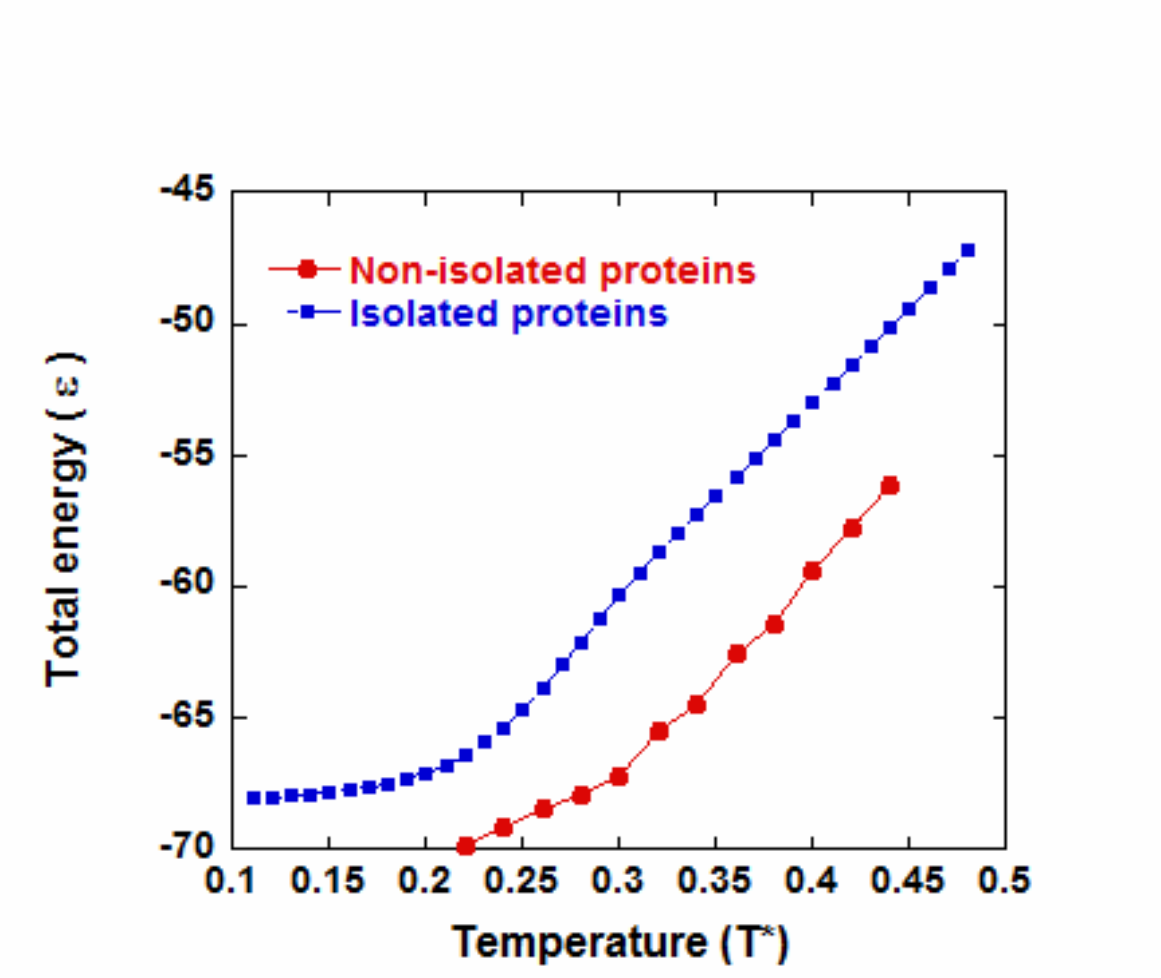


Figure 4.3a: Specific heat curve of the 42mer HP protein as a function of temperature (T*). The peaks at T*=0.49 and T*=0.26 represent the collapse and folding transition of the protein respectively.

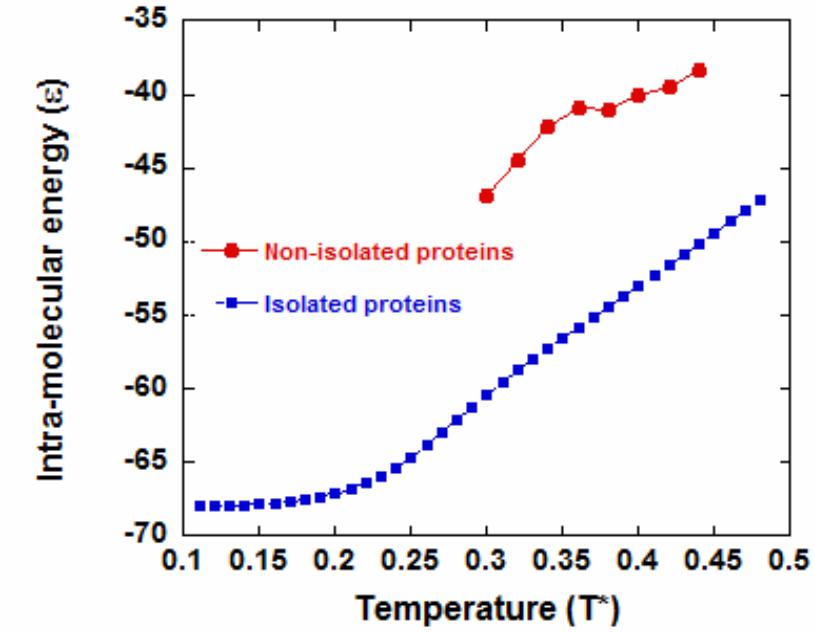


Figure 4.3b: Intramolecular energy of two chains of the 42mer model protein as a function of temperature in the diluted (isolated) and concentrated (non-isolated) regimes

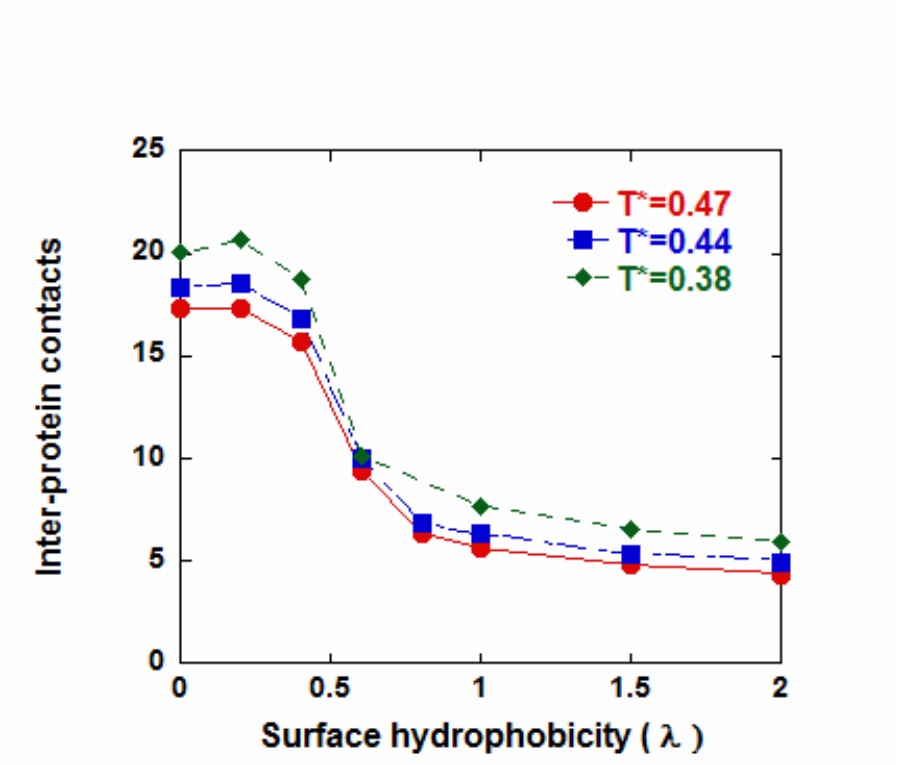


Figure 4.4: Average number of inter-protein contacts as a function of surface hydrophobicity (λ) of two 42mer model proteins at different temperatures (T*)

4.3.2 Protein aggregation on a hydrophobic flat surface

Now we set to discover mechanisms through which protein aggregation can be mitigated. We placed a hydrophobic surface at z=0 and an athermal surface at z_{max} (systems studied with z_{max}=43 and z_{max}=20) and varied the value of λ of the surface at different temperatures. We report four interesting findings.

First, the number of inter-protein contacts decreases with an increase in surface hydrophobicity. Second, there is a drastic decrease in the number of inter-protein contacts at a particular value of λ corresponding to the adsorption of the protein onto the surface. To rationalize these findings

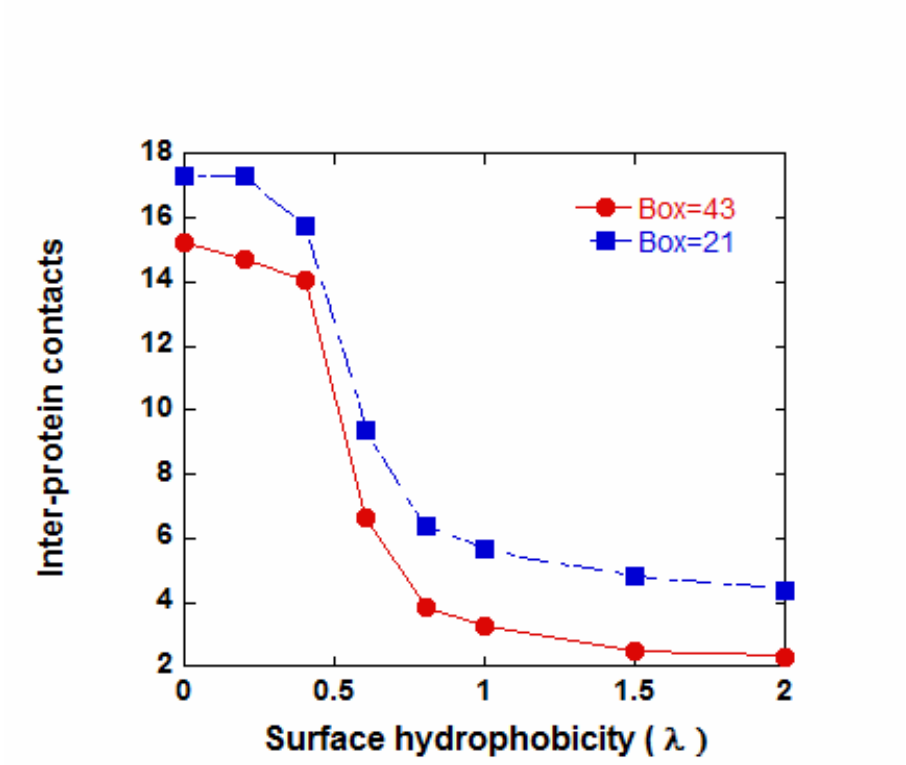


Figure 4.5: Average number of inter-protein contacts of the two 42mer model proteins as a function of surface hydrophobicity at T*=0.47, at two different wall separations.

we note that at very low values of λ, the protein has no propensity to adsorb onto the surface and hence inter-protein attractions in the bulk dominate. For values of $\lambda > \lambda_c$ (λ_c corresponds to the adsorption transition of the protein) the protein surface interaction energy dominates over inter-protein interactions and hence there is a drastic decrease in the number of inter-protein contacts.

Third, the number of inter-protein contacts at any given value of surface hydrophobicity (λ) increases with decrease in temperature. At higher temperatures, the chain entropy is the dominant force and hence there are fewer inter-protein contacts. As the temperature is decreased, energy dominates over entropy and consequently both the inter-protein interaction and the protein-surface interaction increases. (**Figure 4.4**)

Finally, at any given temperature and surface hydrophobicity the number of inter-protein contacts increase as the distance between the confining flat walls decreases, which shows the effect of concentration on protein aggregation. **(Figure 4.5)**

4.3.3 Protein aggregation on negative and positive curvature surface

Now we study the effect of curvature (positive, zero and negative) on protein aggregation. **Figure 4.6** shows the behavior of the 42mer two protein system under different geometries at T*=0.47. The effect of curvature on the percentage loss of the native contacts with varying surface hydrophobicity (λ) at low temperature (near the folding temperature) is shown in **Figure 4.7**. All the above geometries (except radius =4) were considered under the same volume (approximately 8000 lattice sites). The results can be rationalized as below.

a) *Adsorption on the inside (negative curvature) of a cavity*: The adsorption of a protein inside a small athermal cavity (R=4) results in an increased number of inter-protein

contacts compared to that on a flat surface. This is simply a confinement effect, which results in an increased protein concentration. As the value of λ is gradually increased we observe a decrease in inter-protein contacts due to an increase in protein-surface interactions. Because very strong protein surface interactions are necessary to break inter-protein interactions, the percentage loss in the native contacts is highest in the case of this smallest pore (R=4) relative to any other, larger size cavity.

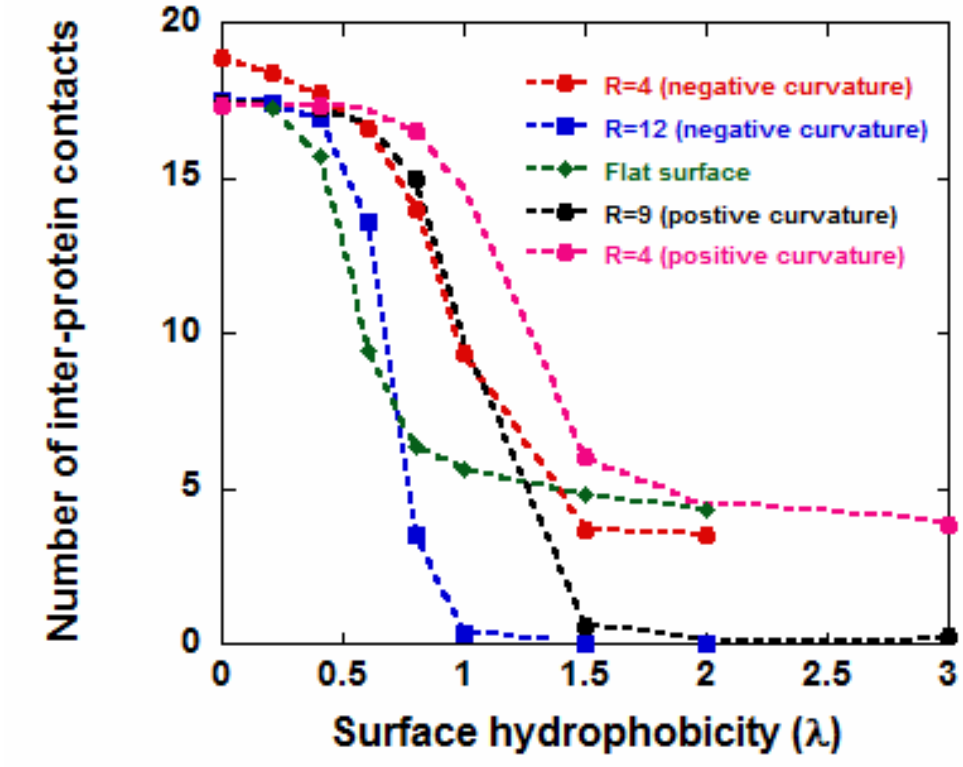


Figure 4.6: Average number of inter-protein contacts of the two 42mer model proteins as a function of surface hydrophobicity (λ) at T*=0.47 in different geometries.

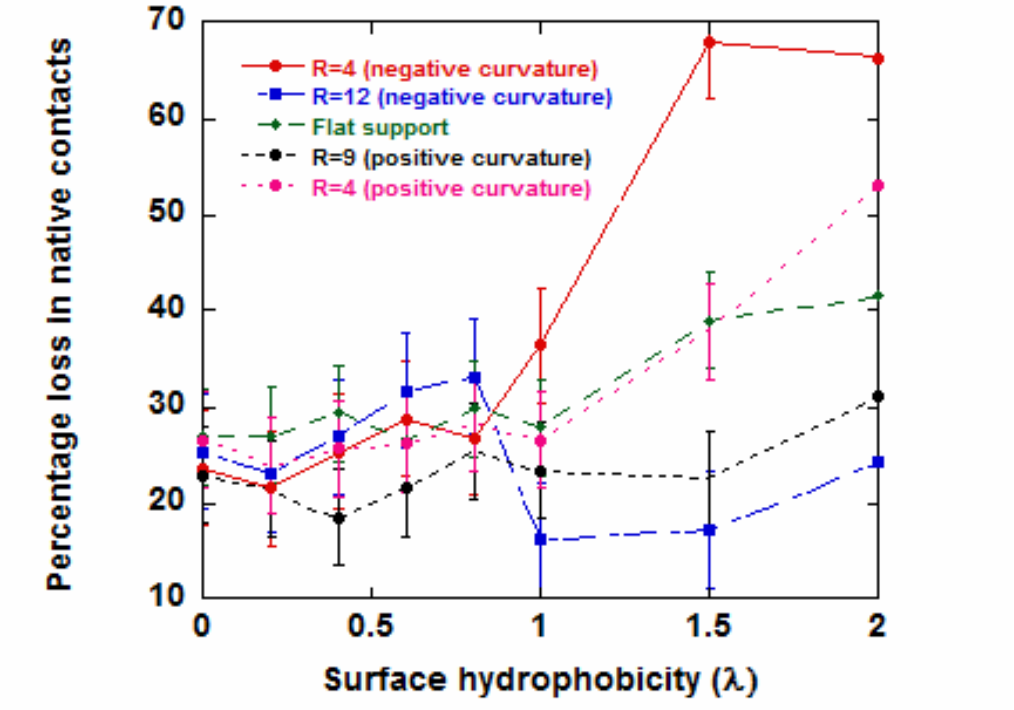


Figure 4.7: Percentage loss in the number of native 'H-H' contacts of the two 42mer model proteins as a function of surface hydrophobicity (λ) in different geometries at T*=0.36

Adsorption inside larger cavities (R=12) proves this point. Since inter-protein interactions are less favored, we do not need such high surface binding to break these inter-molecular associations. Consequently, fewer native contacts are lost relative the R=4 case, especially for intermediate values of λ ($1.0 < \lambda < 1.5$). For higher values of λ ($\lambda > 1.5$), as expected, we lose native contacts due to protein-surface interactions.

Comparing our results for R=4, 12 and a flat surface, we observe that for large enough λ at R=12 there are no inter-protein contacts; the loss of native contacts are also minimized under these conditions. While a more detailed study of different radii would allow us to pinpoint the optimal size more precisely (we have tried different radii upto R=15, but our results are dominated by simulation uncertainties), it is apparent that cavity confinement

has an optimal size where the loss of native contacts are minimized, while inter-protein interactions are completely absent. Presumably, these inter-protein interactions are exchanged in favor of protein-surface interactions. This combined behavior might be the mechanism by which chaperonins disaggregate and then refold proteins.

b) *Adsorption on the outside of a cavity:* An increase in surface hydrophobicity results in decreased inter-protein contacts due to surface adsorption, as discussed above. Adsorption on the outside of small cavities (R=4) cannot completely mitigate inter-protein interactions due to the proximity of the adsorbing protein chains. Instead, adsorption on the outside of larger cavities (R=9) helps to mitigate inter-protein interactions due to increased center to center distance between the adsorbed proteins as shown by previous experiments(122). Note that, while adsorption on the outside seems to track the behavior inside cavities, the behavior is always a little worse. We conjecture that this arises because the threshold energy for adsorption is higher in the "outer" case.

c) To validate that the protein-surface interactions are indeed the major driving force for reduction in inter-protein contacts, the surface energy (protein-surface interaction energy) is plotted in **Figure 4.8** as a function of surface hydrophobicity in different geometries at T*=0.47. In the absence of inter-protein contacts, we expect the adsorption transition to shift from lower to higher values of surface hydrophobicity as we go from negative to positive curvature. Instead, we see a non-monotonic behavior due to the presence of inter-protein interactions. Inside smaller cavities (radius=4) as discussed above, the adsorption transition occurs at a higher value of λ since at these high protein concentration stronger surface interactions are necessary to break the inter-protein contacts. On the other hand adsorption on the outside of cavities (positive curvature)

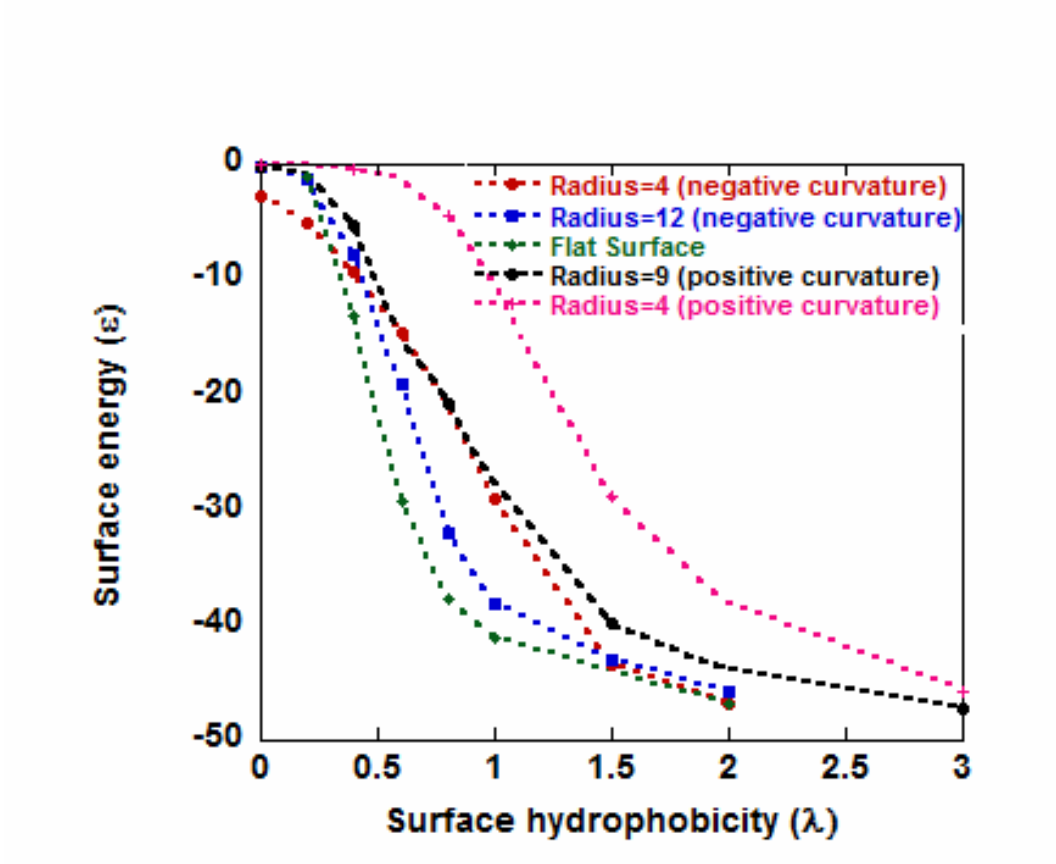


Figure 4.8: Averaged surface energy (protein-surface contact energy) of the two 42mer model proteins as a function of surface hydrophobicity (λ) at T*=0.47 in different geometries

happens at higher values of λ due to weaker protein-surface interactions. An interesting fact emerging from **Figure 4.8** is that, although the surface-protein interaction strength is higher inside the cavity of R=12 than compared to that on a flat surface, the protein-surface contacts are higher on a flat surface because the proteins can expose more surface area for adsorption on a flat surface as compared to that on the inside of a cavity. Many expanded structures are eliminated inside a cavity and consequently some of the 'H' residues now form intra-molecular native contacts. **Figure 4.9** shows the percentage loss of native contacts inside cavities of

different radii at T*=0.36 and λ=1. We see that the optimal loss in native contacts is in case of adsorption inside a cavity with radius 12. Although, increasing the size of the cavity results in lowering the protein concentration, nevertheless it is interesting to see that even at lower concentration (inside R=15), the loss is higher compared to that inside R=12 which emphasizes the role of an optimal curvature

Clearly, adsorption on the inside of cavities of optimal size at moderate hydrophobicity is most efficient in disaggregating proteins and also retaining the native contacts of the protein. These are the main conclusions of our work, which mainly arise due to protein surface interactions which are strong enough to reduce inter-protein interactions but not strong enough to denature the protein. The confinement also provides additional stability by stabilizing the compact folded state due to entropic effects (36, 128).

4.4 Discussion

Although our study provides a rationale to explain the role of chaperonins in protein disaggregation, these conclusions are not general conclusions due to the following caveats. The current simulations were carried out at slightly higher temperatures than the folding temperature of the protein to prevent the system from being trapped in local minima. At low temperatures (T) the simulations cannot overcome the large free energy barriers since the Boltzmann acceptance probability ($\exp(-\Delta U / k_B T)$) is very low; thus our simulations do not allow us to thoroughly sample equilibrium states. Nevertheless this is not far off from the mechanism of chaperonins. The chaperonin (Hsp 70) of the class of heat shock proteins is known to overcome come these

barriers by creating a thermal shock /stress – in our case this is emulated by carrying out the simulations at temperatures higher than the folding temperature.

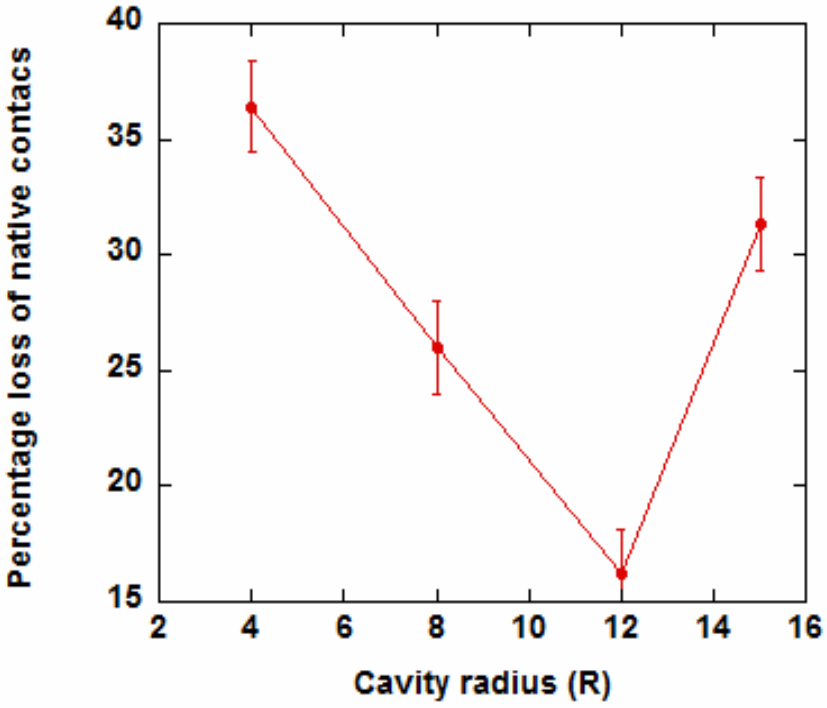


Figure 4.9: Percentage loss in the number of native 'H-H' contacts for the two 42mer proteins inside cavities of different radii at T*=0.36

Despite being a mediator for protein-disaggregation, strong hydrophobic surfaces are also known to denature proteins due to the exposure of the hydrophobic core of the protein. Hence this choice of optimal value of the surface hydrophobicity (λ) where the loss in native contacts is minimum, depends on the proteins studied, as shown by experimental and computational studies

by Zoungrana et al and Zhuang et al. (1) (1) (1) (1) (1) Also hydrophobic attractions between the 'H' residues of the protein should be weakened near a hydrophobic surface compared to that in the bulk, as shown by computational works of Patel et al (130) which is not captured by the HP model. Consequently the model overestimates the value of λ required to disaggregate proteins under different geometries**.**

4.5 Conclusions

In the current study, we show how protein disaggregation can be mediated through strong protein surface interactions. We further show that disaggregation is most efficient in the case of adsorption inside cavities (negative curvature) of optimal size compared to that on flat supports and on surfaces with positive curvature. Besides, adsorption inside such cavities helps to regain the native contacts which were lost due to inter-protein interactions. We believe our work provides a basis to explain why enzymes immobilized in porous media retain more activity than in free solution and also to the role of chaperonins in protein disaggregation and folding.

Chapter 5

Enhanced Wang Landau Sampling of Adsorbed Protein Conformations

Using computer simulations to model the folding of proteins into their native states is computationally expensive due to the extraordinarily low degeneracy of the ground state. In this chapter, we develop an efficient way to sample these folded conformations using Wang Landau sampling coupled with the configurational bias method (which uses an unphysical "temperature" that lies between the collapse and folding transition temperatures of the protein). This method speeds up the folding process by roughly an order of magnitude over existing algorithms for the sequences studied. We apply this method to study the adsorption of intrinsically disordered HP protein fragments on a hydrophobic surface. We find that these fragments, which are unstructured in the bulk, acquire secondary structure upon adsorption onto a strong hydrophobic surface. Apparently, the presence of a hydrophobic surface allows these random coil fragments to fold by providing hydrophobic contacts that were lost in protein fragmentation.

5.1 Motivation

Computer simulations have been extensively used to understand the physics of protein folding (131). Ideally, one would like to model the protein and the surrounding solvent in full atomistic detail; however, such calculations are very intensive to carry out even on the fastest computers (132). Thus, simpler coarse-grained models like the HP model (110), Go model and others (133) are employed to gain important physical insights into the physics of these fascinating molecules. It has been shown that the folding of a HP protein is NP complete– thus sampling all possible configurations would require an amount of time that varies roughly

exponentially with the size of the protein (51, 112).Advanced simulation techniques, like configurational bias (113) and umbrella sampling (134) have been used to circumvent this crucial computational bottleneck. Multiple Histogram (114, 115) methods have been employed to get better estimates of the thermodynamic quantities from simulations performed at different conditions. The accuracy of these techniques deteriorates near the melting temperature of a protein, wherein the system takes progressively longer to relax. At temperatures lower than the melting temperature, it becomes difficult to cross the free energy barriers separating the folded and unfolded states, and hence the sampling efficiency is poor.

Wang Landau (WL) sampling (135) has proven to be quite effective to counter such sampling problems by directly calculating the density of states (DOS) of polymers and simple protein models. WL sampling is successful in sampling across the free energy barriers because in this method a configuration in an energy state is not weighed by the Boltzmann factor of its free energy but by the inverse of the density of that energy state. Hence, WL sampling attempts to perform a uniform sampling of all energy states. Since the DOS of the system is not known *apriori*, WL sampling starts with an initial guess for the DOS, and keeps updating it on the fly. Upon convergence, a good estimate of the DOS is thus obtained. However, the WL sampling becomes inefficient for systems in which some energy states have low degeneracy, that is, there exist energy states which can only be obtained through very specific configurations. Proteins, the molecules with a unique folded conformation, come under this special category, wherein the WL sampling is not able to efficiently sample this folded state. More efficient Monte Carlo (MC) moves, such as, pull moves and bond re-bridging moves, do enhance the sampling efficiency but at the expense of extremely large computational times (136). The idea of combining Wang

Landau sampling with configurational bias was introduced in the work by Jain et al (137). In this work, Wang Landau sampling was coupled with configurational bias carried out at infinite temperature to increase the efficiency of the algorithm. In subsequent works by Rathore et al Wang Landau sampling was carried out in combination with Molecular dynamics, Monte Carlo and Parallel tempering simulations (138-140). One way to increase the efficiency of WL sampling to ensure proper sampling of low energy states would be to introduce an energetic bias in the step that generates new configurations. This approach was used by Castells et al (141).in which WL sampling was done with configurations generated by an energy bias using a local definition of temperature. They defined local temperature as the energy derivative of the DOS of the present energy state, *Ω(E)*,

$$\beta = \frac{1}{k_B T} = \frac{\partial(\ln(\Omega(E))}{\partial E} \qquad (1)$$

However, since *Ω(E)* is not known accurately, this "temperature" can fluctuate strongly, and can even go negative. This procedure, though interesting, does not yield an improved algorithm. In this paper we show that the WL algorithm, when coupled with a generation step which uses an energetic bias at a constant pseudo temperature, yields an efficient sampling method for protein systems. Further, we show that the sampling efficiency is maximized when the pseudo temperature is selected between the collapse and folding transition of the protein. The method was tested for benchmark HP sequences published elsewhere and was able to reproduce the results. For the HP model studied by Li et al (142), we were able to get the density of states of their 36mer adsorbing on a weak hydrophobic surface within 28 hours on an Intel Xeon 2.8.GHz machine. (The authors claim that it took them more than 10days on a AMD 2.2GHz dual core

processor). Thus, for the above system we were able to get roughly an order of magnitude improved over the WL simulation protocol in the above paper.

An understanding of conformational changes in proteins upon adsorption has been a topic of great interest and debate (143-145). Recent studies have shown that adsorption of proteins on hydrophobic surfaces can enhance their structure and stability (108). In this paper, we apply our method to study the adsorption behavior of unstructured protein fragments on a hydrophobic surface. We observe that the protein fragments, which remain unstructured in the bulk environment, attain secondary structure upon adsorption. Similar findings were recently published by Shea et al (1).

5.2 Model

We studied the Hydrophobic (H) –Polar (P) lattice protein model, commonly known as the HP model, introduced by Dill (110). In this model, amino acids are represented by beads labeled as either hydrophobic (H) or polar (P). The self-avoiding chain representing a protein molecule resides on a three dimensional cubic lattice. The bond length between any two amino acid beads is kept constant equal to the lattice constant. Solvent molecules are accounted for in an implicit manner by introducing an attraction between two H beads to account for the solvent-driven hydrophobic effect. Energetic interactions between the H beads are considered to be $-\varepsilon$, when the two H groups are nearest non-bonded neighbors and zero otherwise. All other interactions (H, P and P, P) are set to zero. Although this model is a simplistic representation of real proteins, it captures many aspects of the physics of protein folding (111, 146-148). In this paper, we have studied a class of truncated peptides which are unstructured in the bulk - a β-

hairpin model and a helix- coil- helix model introduced by Bellesia et al.(149) based on the design of proteins by Hetch and co workers (5). The performance of the algorithm is studied using a 36mer HP protein.

5.3 Simulation Details

Simulations are performed in a cubic box with lateral size greater than the length of the chain studied. This size of the simulation box is large enough to avoid possibility of any interactions between the protein chain with its periodic image. Periodic boundary conditions are imposed in the x and y directions. Two surfaces, each modeled as a two-dimensional plane, are placed at z=0 and z=z_{max}. The interaction energy between the plane at z=0, termed as the hydrophobic surface, and a H bead of the protein is -$\lambda\varepsilon$ when the H bead is at the z=1 plane, and zero otherwise (λ=0 represents a non-interacting or an athermal surface and λ=1 represents a strongly hydrophobic surface). For the surface at z_{max}, λ=0. The total energy of the protein is the sum of intramolecular H-H contacts and H-surface contacts at z=0 plane ($E_{total} = E_{intra} + E_{surface}$).

Table 5.1 lists the sequences used in this study. **Table 5.2** lists the ground state energy (total energy) of the models studied in bulk and upon adsorption onto a hydrophobic surface. WL sampling with configurational bias is performed on these model peptides to obtain the DOS of the system. Thermal and structural stability are inferred from the specific heat curves.

Table 5.1- List of sequences studied

Sequence	
Helix-coil-helix	HHPPHHPPHHPPHHPPHHPPHHPPPPPPPHHPPHHPPHHPPHHPPHHPPHHP PHH

β-hairpin	HPHPHPHPHPHPHPHPPPPPPHPHPHPHPHPHPHPH
36mer	PPPHHPPHHPPPPPHHHHHHHHPPHHPPPPHHPPHPP

Table 5.2 Ground state energy (total energy) of the models studied

Model	Ground state energy	
	Bulk	Adsorbed
Helix- coil – Helix	-34	-50
β- hairpin model	-21	-31
36mer	-18	-241 (weak attractive surface)

5.4 Algorithm

In WL sampling, an energy state with energy E is sampled with a probability proportional to the reciprocal of its density of state, $\left(\frac{1}{\Omega(E)}\right)$. The value of $\Omega(E)$ is not known *a priori* and hence an initial guess for $\Omega(E)$ is taken as the starting point. WL sampling is started from a random state with energy E_1 from which a new state E_2 is generated and accepted with probability min ($1,\Omega(E_1)/\Omega(E_2)$). Each time an energy state, E is visited, the $\Omega(E)$ is updated by $\Omega(E) = \Omega(E) \times f$, where f is called modification factor ($f>1$). Along with this, a counter or a

histogram, *H(E)* is maintained for each *E*, which is incremented by 1 each time a state is visited. WL sampling is performed until we have a sufficiently flat histogram i.e., $H(E) \geq p \times \langle H(E) \rangle$ for all *E*, where $\langle H(E) \rangle$ is the average of all histogram values, *p* is the flatness criterion. Once we have a flat histogram, all histogram values are reset to zero, the modification factor f is changed to $f_{new} = f^{a}$, where $0 < a < 1$, and the random walk is again started from a randomly selected configuration. The WL simulation is continued in the above described manner till *f* becomes less than a value, f_o. For our simulations, we have set the initial guess for $\Omega(E) = 1$ for all *E*; $p = 0.8$; $a = 0.5$ and $f_o = \exp(10^{-8})$. The above is a basic description of the WL sampling algorithm. In the present study, we use the configurational bias method to sample the low energy states efficiently. This involves randomly selecting a segment *i* from a chain of *N* segments and subsequently removing either segments *i* to *N* or *1* to *i*, with equal probability and then re-growing the chain with a bias towards lower energy conformations. This bias is removed during acceptance. A segment *i*, can be grown in *k* different directions. The energy of a trial direction *j* is denoted by $u_i(j)$. From the *k* possible directions, we select one, say *n*, with a probability $p_i(n) = \frac{\exp(-\beta u_i(n))}{w_i(n)}$, where $w_i(n)$ is defined as $w_i(n) = \sum_{j=1}^{k} \exp(-\beta u_i(j))$. This step is repeated until the entire chain is grown and the Rosenbluth factor *W(new)* for the new chain is calculated. *W(new)* is defined as $W(new) = \prod_{i}^{N} w_i(n)$. The old chain is retraced and *W(old)* is calculated. This chain regrowth move is accepted with a probability

$$\min\left(1, \frac{\Omega(E_{old})W(new)\exp(\beta(E_{new} - E_{old}))}{\Omega(E_{new})W(old)}\right) \qquad (2)$$

It can be shown that using eqn(2) as the acceptance probability satisfies detailed balance to the approximation of *ln f,* similar to the original WL algorithm. On the contrary, it can be shown that the algorithm proposed by Castells et al, does not satisfy detailed balance. The novelty of our method is the way in which the growth step of the chain is biased to enhance sampling of low energy states by assuming a pseudo temperature, β^{-1}, preferably between the collapse and folding temperatures of the protein. Our method shows a significant enhancement in sampling the DOS of the sequences listed in **Table 5.1** over the original WL algorithm. The improvement in computational efficiency from our algorithm over the original WL algorithm will vary depending on the studied sequence. In the case of the 36-mer, we found roughly an order of magnitude increase in computational efficiency from our method as compared to the computational effort reported in the Li et al. paper, even though in their work efficient MC moves such as pull moves and bond re-bridging moves were also employed. WL sampling does a random walk in the energy space at infinite temperature and hence takes much longer time to explore these low energy states which are observed near the folding temperature. While very low values of β would not efficiently bias the generated configurations towards low energy states, a very high value of β would also be inefficient as the system may get trapped only in low energy states. Based on our observation, we find that the optimum convergence rate is obtained when the value of the pseudo temperature (β^{-1}) is between the collapse and folding transition temperatures of the protein.

5.5 Results and Discussion

The performance of our algorithm is tested for the adsorption of a 36mer HP protein on a weak hydrophobic surface. This system has been previously studied by Li et al. who have

reported the time required to establish the density of states of the system as more than 10 days. **Figure 5.1** shows the specific heat, C_v curves for adsorption of a 36mer HP protein on a weak hydrophobic surface ($\varepsilon_{H\text{-}H}$=12, $\varepsilon_{H\text{-}Surface}$=1) at different values of pseudo inverse temperature, β and with the reference data. The C_v curves overlap with each other within the error bars enforcing the fact that the choice of β does not affect the final value of DOS of the system.

Table 5.3 shows the simulation times required for convergence of our algorithm for different values of β. We see that for the values of β^{-1} between the folding and the collapse temperature, the computational time is least. The computation time does not vary significantly for the values of β between 0.16 and 0.33. Therefore, an accurate determination of β is not necessary for our method to give computational improvement over the traditional WL algorithm. The computational time increases for either very high (~ 1.0) or very low values of β (~0.04), but regardless of this, our method appears more efficient than the WL algorithm. The computation time for the optimum value of β (~24hr) is roughly an order of magnitude less than the time quoted in Li et. al (> 240 hr) for the 36-mer. **Figure 5.2** shows the computation time required to establish the DOS of the system at different values of β. The simulation times are also listed in **Table 5.3**. Although we do not have a systematic recipe for determining the optimum value of β, but since the computational efficiency does not vary much if the value of β^{-1} is between the collapse and folding temperature of the protein, an estimate of the folding and the collapse temperatures should be suffice in most cases. The collapse transition of a protein or a polymer can be obtained by doing a WL sampling only for a high energy window. This simulation converges very rapidly and gives a good estimate of the collapse temperature. To illustrate, we applied the above described approach to determine the collapse transition temperature for the 36-

mer protein near a weak hydrophobic surface, WL sampling with configurational bias (with β=0) was carried out in the energy window 0 to -200 (the ground state energy of the system is -241. The DOS converged rapidly within 30 mins. The collapse temperature was extracted from the specific heat curve. This temperature, when used as the pseudo temperature, gave accurate density of states of the entire system. The actual value of collapse temperature is T*= 5.93 and the value of collapse temperature got by sampling only at higher energy window is T*= 6.37.

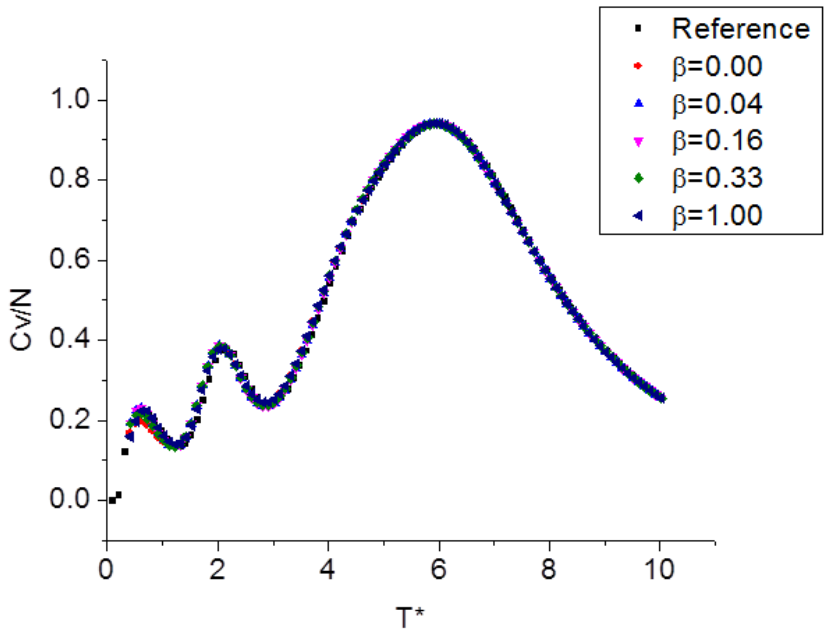


Figure 5.1: Specific heat curve of adsorption of 36mer HP protein on a weak hydrophobic surface ($\varepsilon_{H\text{-}H}=12$, $\varepsilon_{H\text{-surface}}=1$) at different pseudo inverse temperature β.

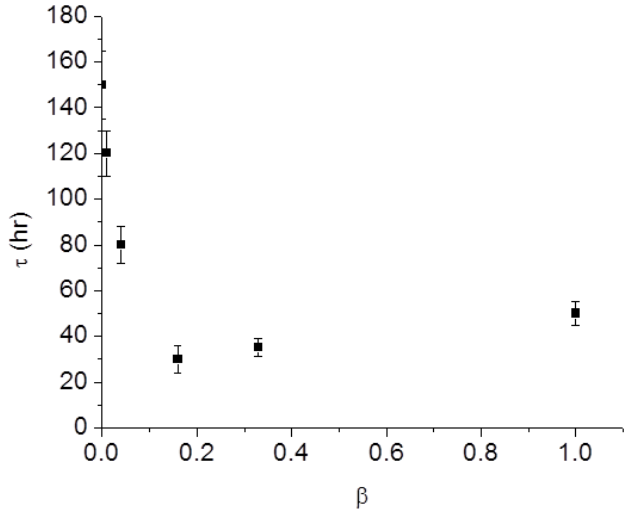


Figure 5.2: Computation time required for convergence for different pseudo inverse temperature of 36mer HP protein adsorption on a weak hydrophobic surface. (ε H-H=12, ε H-surface=1)

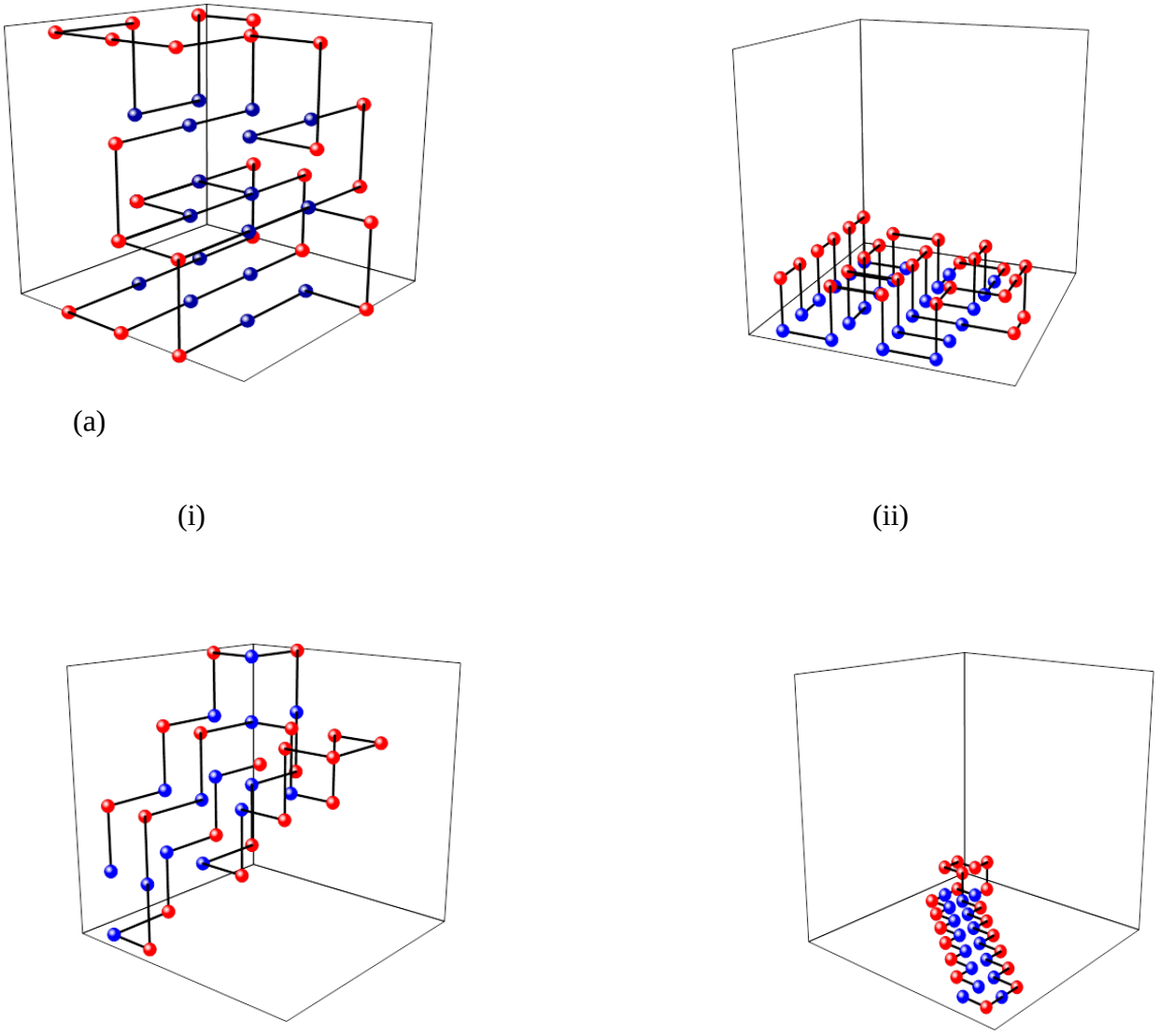
Table 5.3: Convergence time of the Wang Landau simulation with configurational bias as a function of the parameter β 9inverse of the temperature)

β (1/T*)	τ (hrs)
0.0	150±15
0.01	120±10
0.04	80±8
0.16	30±6
0.33	35±4
1.00	50±5

Our method is applied to study the adsorption of unstructured peptides on a strong hydrophobic surface (λ=1). A pseudo temperature of β^{-1}=1.00 was used in the simulations for these models. **Figure 5.3a** and **5.3b** show the ground state structure of a helix-coil-helix and β-hairpin model peptide, respectively, in the bulk and when absorbed on a strong hydrophobic surface. In the bulk, the helix-coil-helix motif is unstructured, while upon adsorption we see induction of helical domains. The β-hairpin protein model is slightly structured in the bulk, but forms β sheets on adsorption. In the bulk, these peptides form a ground state structure with a hydrophobic core and polar exterior to maximize H-H contacts which prevents them from attaining an ordered structure. On a hydrophobic surface, the peptide exposes its H groups to the surface. Though a small number of internal contacts are lost, the peptide regains this energy by interacting with the surface. This explains the formation of domains in proteins, as has also been discussed elsewhere

Figure 5.4a and **5.4b** show the specific heat of the helix-coil-helix and the β-hairpin model, respectively. In the bulk, the helix-coil-helix peptide only has a collapse transition at T*=0.57 (**Figure 5.4a**). Upon adsorption onto a strong hydrophobic surface, the peptide exhibits two additional distinct transitions than in bulk. The specific heat is characterized by an adsorption transition at T*=1.06 and a collapsed transition at T*=0.62 and the shoulder at T*=0.20 is believed to be a result of structural rearrangements on the surface. The β-hairpin peptide in the bulk is mainly characterized by a folding transition at T*=0.40 (Fig. 4b). Upon adsorption onto a hydrophobic surface, we see an adsorption transition at T*=0.91 and a folding transition indicated by the shoulder at lower temperatures. In the case of both peptide models, the collapse transition is shifted to higher temperatures indicating higher thermal stability of the adsorbed peptide. Adsorption experiments have reported that amyloid β-peptide, which is disordered in the

bulk, attains an ordered structure upon adsorption. Enhancement in structure and stability of proteins has also been predicted via simulations.

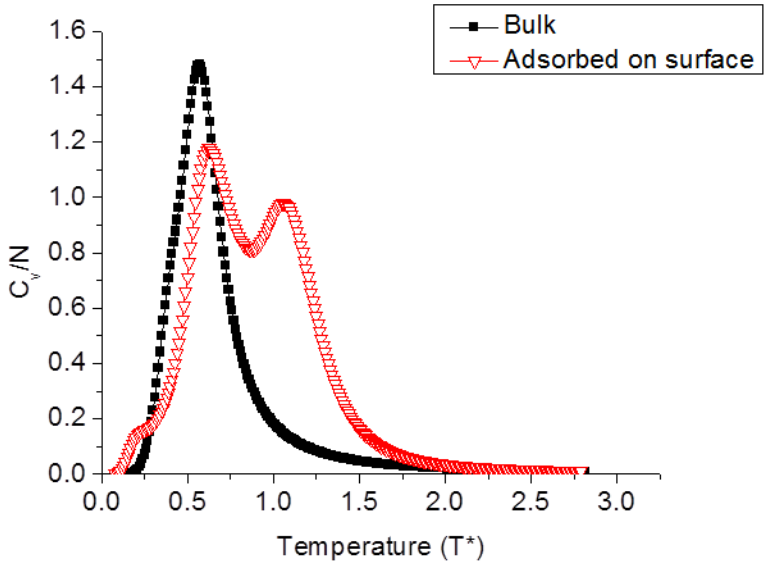


(a)

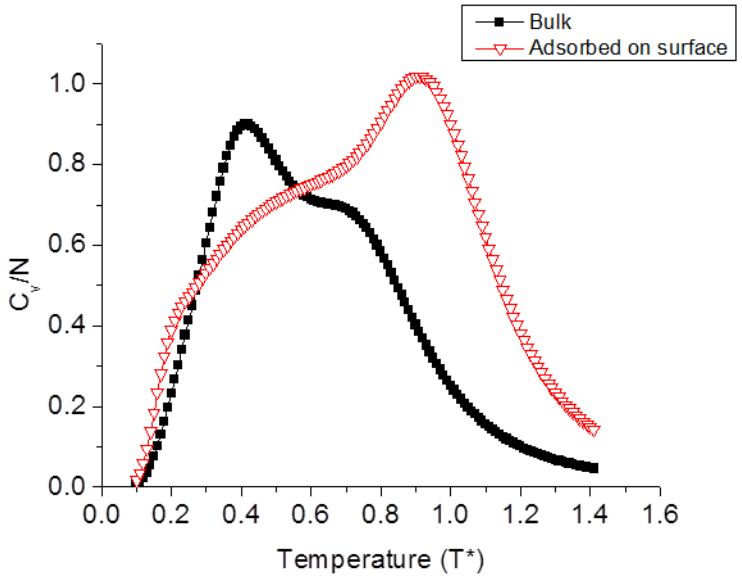
(i) (ii)

(b)

Figure 5.3: a) Structure of helix–coil-helix model peptide i) minimum energy state in bulk ii) minimum energy state upon adsorption (λ=1.0) b) β-hairpin model peptide i) minimum energy state in bulk ii) minimum energy state upon adsorption (λ=1.0) . H groups are represented in blue; 'P' groups are represented in red



(a)



(b)

Figure 5.4: a) Specific heat of helix-coil-helix model peptide and b) β-hairpin model peptide in bulk and when adsorbed onto strong hydrophobic surface (λ=1)

5.6 Conclusions

We have shown that Wang Landau sampling coupled with the configurational bias method samples low energy states roughly an order of magnitude faster than the current best methods. This allows us to model the folding of proteins in a tractable and efficient manner. The optimum pseudo temperature for most efficient WL sampling is found to be between the collapse and folding temperature of the protein. Using this novel method we studied the adsorption of unstructured peptides on a strong hydrophobic surface. We show that these protein fragments ~~can~~ gain structural and thermal stability on adsorption onto such a surface. This explains experimental findings which suggest that, though individual domains of proteins may not be structurally stable, but when two domains come together, they complement each other to form a stable tertiary structure. (107)

Chapter 6

Design of protein sequences using explicit solvent model

In my previous work, I studied the stability of proteins using an implicit solvent Hydrophobic – Polar (HP) lattice model. Although, helpful in understanding the physics of protein folding, it is very simplistic in nature. The atoms are constrained to reside in a lattice and therefore lack conformational degrees of freedom and most importantly due to the absence of solvent, the temperature dependence of hydrophobic effect is not captured. In this chapter, I have tried overcome this limitation by modeling proteins using an explicit solvent model.

6.1 Introduction

Water is a natural solvent. The vast majority of interactions in the biological systems take place in an aqueous environment. It has long been appreciated that water is complex and exhibits some unusual if not unique characteristics. Liquid water exhibits a maximum in density and isothermal compressibility as a function of temperature. It solidifies with volume increasing at low pressure and the solid phase exhibits a variety of crystalline structures. Although, these properties of water have been known since long time, the origin for this anomalous behavior is still controversial (150, 151). Due to the electronegativity difference between the hydrogen and oxygen atoms, they develop a partial charge which drives the formation of hydrogen bonding between the hydrogen and oxygen atoms of the neighboring molecule at low temperature (152, 153).Numerical simulations based on some of the available pair potentials for the interaction between the water molecules like TIP3,TIP4 and SPCE reproduce many of the properties quite well but are limited due to computational constrains (154, 155). In the current study, we want to

choose a model for water that captures all the anomalies and is simultaneously computationally economical.

6.2 Jagla Potential

Jagla potential has shown to mimic most of the properties of waster. Water molecules are modeled as spherical particles interacting though a spherically symmetric potential. The potential is characterized by a hard core potential, a repulsive ramp and followed by an attractive tail. The attractive tail is added to capture the liquid gas transition. The model is capable of explaining the anomalies even in the absence of an attractive tail (156-158). The interaction potential U(r) of Jagla solvent particles is characterized by five parameters. The hard core diameter **a**, the soft core diameter **b**, the range of attractive interaction **c**, the well depth of attraction $\mathbf{U_A}$ and the height of the repulsive ramp $\mathbf{U_R}$ as shown in **Figure 6.1**. There five parameters can be reduced into three independent parameters **b/a, c/a** and $\mathbf{U_R/U_A}$. In our study we have set **b/a=1.72, c/a=3.00,** $\mathbf{U_R/U_A}$ **=3.5** (159-162). These values are chosen taking into account the distances between the central water molecule to the first and second neighbor shells as measured by the distance between the oxygen-oxygen atoms in the radial distribution function of water molecule. This choice of parameters reproduces the phase diagram of water, with two critical points, one at high temperature corresponding to the liquid-gas transition and other at low temperature corresponding to the liquid-liquid transition. The two length scale potential captures the properties of water. At high pressures, in order to minimize the free energy, volume is minimized whereas at low pressures energy is minimized. This is to emulate the property of water that at low temperatures, water forms hydrogen bonds whose bond distance is comparatively more than at high temperature when these bonds are broken (161-163).

Monte Carlo simulations were carried out at constant pressure at both in the high pressure regime (P*=3.0) and low pressure regime (P*=0.05) at a temperature of T=1.0. The system is equilibrated and probability distribution of the pair distances between the Jagla particles is calculated.

Figure 6.2 and **Figure 6.3** show the radial distribution function of a system of spherical particles interacting according to the Jagla Potential. As discussed before, at low pressure (P*=0.3) we see a sharp peak at r=1.72 corresponding to **b** and a high pressure (P=3.0) this peak shifts to r=1.0 corresponding to **a**.

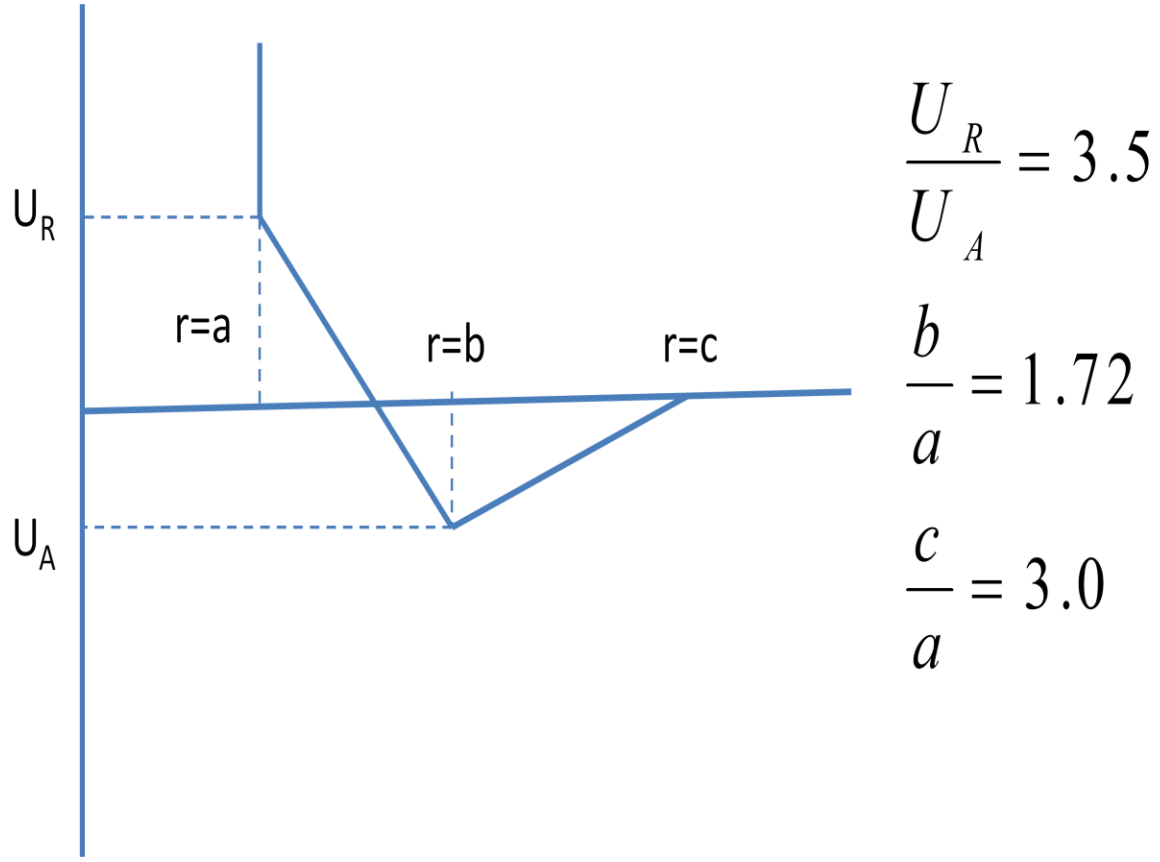


Figure 6.1: Schematic representation of Jagla potential

6.3 Modeling Hydrophobic solutes

The solubility of non-polar compounds in water is very low. A notable feature of the hydrophobicity is that the solubility decreases with increasing temperature at low temperature ie the hydrophobes become more hydrophobic (164-166). At high temperatures, the solubility increases monotonically with temperature. This means the solubility has to go through a minimum as a function of temperature. The low solubility is as a result of unfavorable solvation energy and a decrease in solubility with temperature is as a result of favorable solvation enthalpy. At low

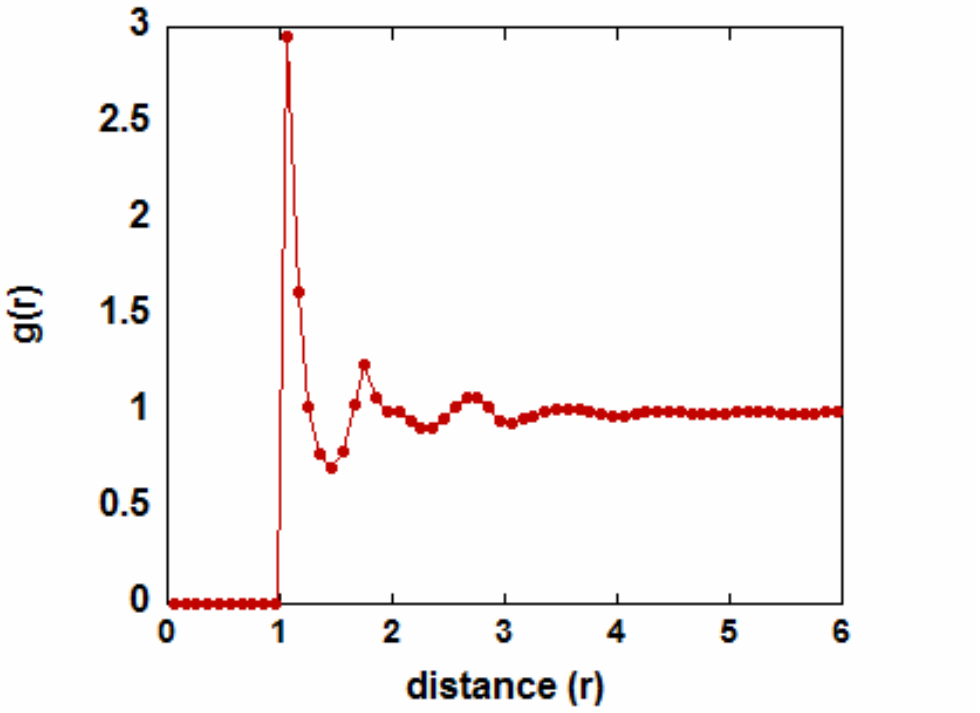


Figure 6.2: Radial distribution function of a system of spherical particles interacting through Jagla Potential at a pressure of P*=3.0 at temperature T*=1.0

temperatures, the increased solubility is as a result of the fact the hydrophobes can occupy the interstitial sites in the water molecule. At high temperatures, the mixing entropy dominates and consequently the solubility is higher. This also provides us a basis to explain protein denaturation at high and low temperatures. Since the solubility of hydrophobes increases at low temperature, the hydrophobic groups which form the core in the active state of a protein is exposed. This structural disruption leads to protein denaturation (162, 165, 167, 168).

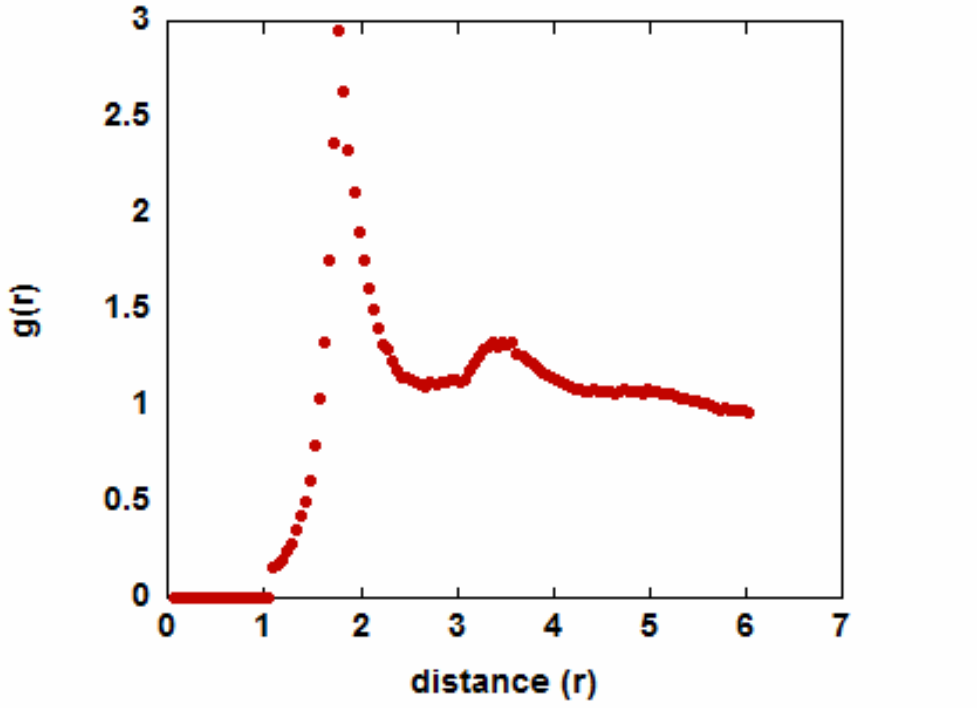


Figure 6.3: Radial distribution function of a system of spherical particles interacting through Jagla Potential at a pressure of P*=0.05 at temperature T*=1.0

Buldyrev et al (162) showed that spherical particles modeled through hard-sphere potential dissolved in water like Jagla solvent mimic the characteristics of hydrophobic solutes in water. They exhibit a minimum in solubility as a function of temperature and further, hydrophobic chains modeled using hard-sphere potential exist in a collapsed state only at intermediate temperatures corresponding to the biological temperatures. At lower temperatures, they become well solvated due to favorable enthalpy and at higher temperature due to favorable conformational entropy.

6.4 Protein Adsorption on hydrophobic surfaces

Hard-sphere particles in Jagla solvent behave similar to hydrophobic solutes in water as discussed above. In the same vein, to study protein adsorption on hydrophobic surfaces, we model the hydrophobic surface as a hard-wall placed at z=0 and $z=z_{max}$. Periodic boundary conditions were imposed in the x and y directions. The effect of the hydrophobic surface on a binary mixture of hard-sphere solute and Jagla solvent is studied through Canonical Monte Carlo simulations carried out at a temperature T*=1.0 and at various number densities. **Figure 6.4** shows the distribution of hard-sphere solutes as a function of distance from the surface. The profile is symmetric since we have placed the surface at both ends of the simulation box. We see an enrichment of the hard-sphere particles and a depletion of Jagla solvent near the surface. The enrichment gets higher when we move from lower to higher solvent number density. This enrichment can be attributed to the entropic effect. Since, hard sphere particles have no interaction until contact; the system free energy is minimized when the entropy is maximized. We know that a particle near the surface excludes less volume compared to the particle in the bulk. Consequently, hard-sphere particles are preferentially adsorbed on the hydrophobic surface

and the Jagla-particles are expelled from the surface. These results provide us a rationale for experimental observation of expulsion of the solvent near the hydrophobic surface during protein adsorption.

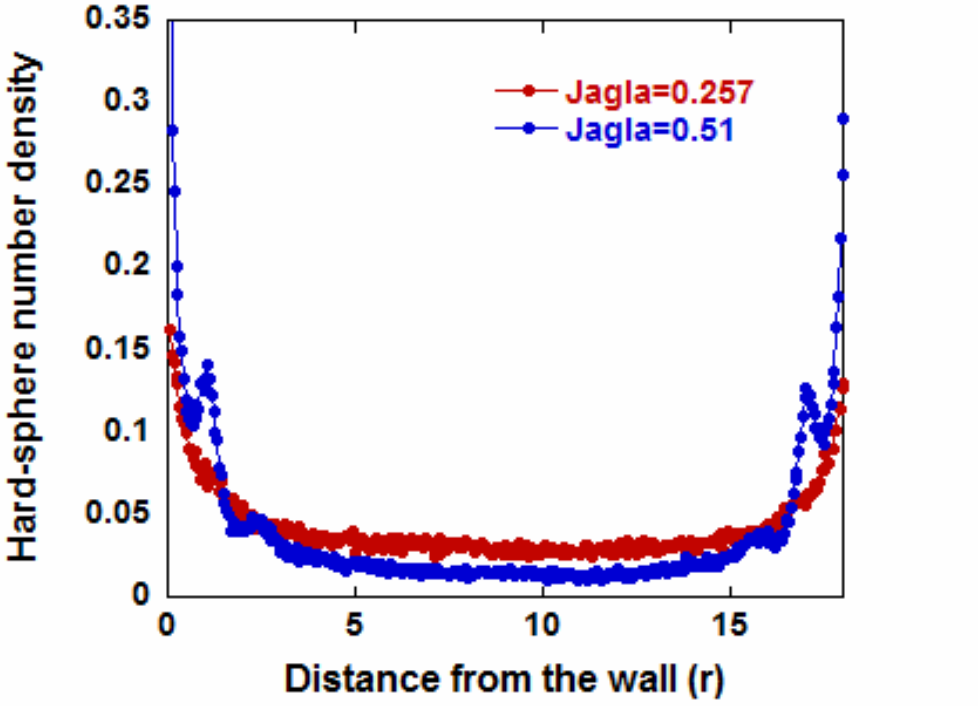


Figure 6.4: Variation in the number density of hard-sphere as a function of distance from the hydrophobic surface. (Hydrophobic surface is located at r=0 and r=18) at different solvent (Jagla) solvent.

Now we study the behavior of short polymer hydrophobic chain in a Jagla solvent (162). The hydrophobic chain is modeled using hard-spheres of the same diameter as the solvent molecule name $\mathbf{d_0 = a}$. We model covalent bonds by linking the beads using a simple potential given below

$$U_{bond}(r) = \begin{cases} \infty & (r < d_1) \\ 0 & (d_1 < r < d_2) \\ \infty & (r > d_2) \end{cases}$$

In the current study we have chosen $d_1 = d_0 = a$ and $d_2 = 1.2d_0$. Canonical Monte Carlo simulations were carried out at a temperature T*=1.0 and a solvent density 0.30. Two interesting findings emerge from the simulations. Firstly, as seen in the case of hydrophobic solutes, the Jagla solvent is expelled from the surface and the hydrophobic chain is adsorbed onto the surface. Secondly, the radius of gyration tensor calculated shows that the chain is completely stretched in the xy direction indicated by the magnitude of the $\mathbf{R_{xx}}$ and $\mathbf{R_{yy}}$ components of the radius of gyration tensor. In summary, a hard-sphere chain in a Jagla solvent adsorbs onto the hydrophobic surface by expelling the Jagla solvent from the surface which is purely driven by entropic effects.

$$R_g^{\,2}(T^* = 1.00) = \begin{pmatrix} 7.061 & 0.326 & 0.0032 \\ 0.326 & 9.61 & 0.0402 \\ 0.0032 & 0.0402 & 0.91 \end{pmatrix}$$

6.5 Protein sequence design using hard-sphere and Jagla potential

The folding of proteins into helices and beta sheets has been investigated for the past five decades (169-175). To date, there has not been a thorough understanding of the physics behind this. Theoretical predictions by Zimm and Bragg (171, 176) give us some information but many studies have revealed contradicting results. For example short chain polyalanine, which according to Zimm and Bragg should form a random coil forms an unusually stable helix in solution and some peptides which have a low propensity for helix formation according to the

theory tend to form stable helices (174, 177). In the current work we seek to explain the helix and the beta sheet forming properties of individual short peptides using a simplistic model with explicit solvent through Monte Carlo simulations. The protein back bone and side chains are represented as either hard spheres (hydrophobic) or Jagla particles (hydrophilic) in a Jagla solvent. The helix and beta sheet forming tendencies are studied while varying the chain length, backbone sequence, size of the side group and the strength of the interactions.

Hydrogen bonding between the amino acid residues is believed to the major driving force behind protein molecules acquiring secondary structures like α-helix and β-sheets. Note however, the role of entropic gain when the solvent is expelled from the hydrophobic core is also known to be a factor driving proteins to fold into secondary structures. In the previous work from our group, we developed a strategy to design protein sequences using hard-sphere and Jagla particles modeled as hydrophobic and polar amino acid residues respectively that collapse into structures with a hydrophobic core and a polar exterior. The study reiterated the fact that, it's the sequence

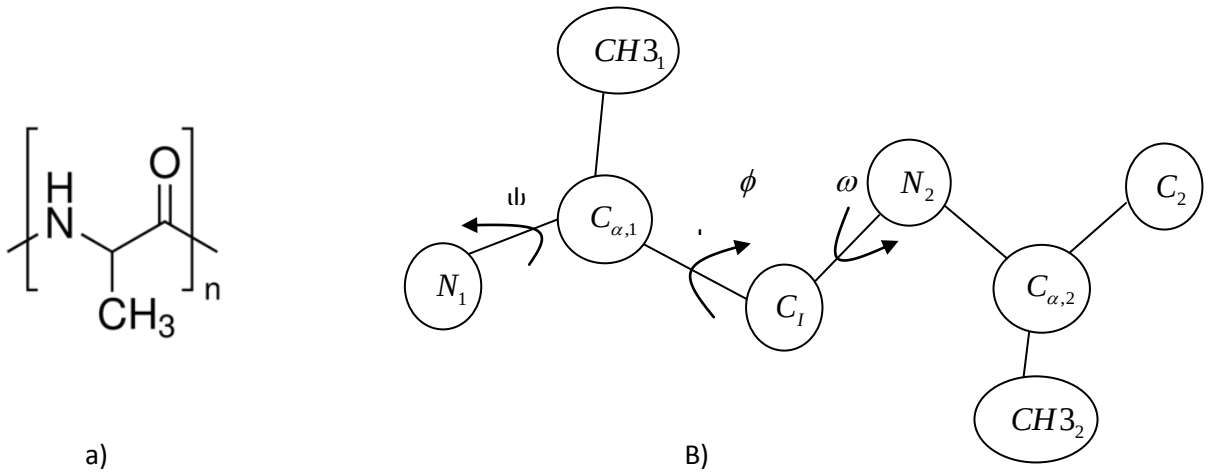


Figure 6.5: a) Chemical Structure of Polyalanine b) Four bead model for Polyalanine

of hydrophobic-polar residues in a protein that determines its secondary structure upon folding (178). In my current work, I wanted to expand on the above to design protein sequences which acquire secondary structures. I started my design, using the simplest amino-acid Alanine as the prototype.

6.6 Polyalanine

The structure of Alanine is shown in **Figure 6.5.** Inorder to model this, the side chain methyl group (CH_3) is modeled as hard sphere particle, N-H and C=O group by Jagla particle and C_α-H group by a hard sphere. Simulations were carried out for different chain length with the amino acid residues in the chain varying from 10 to 50. A single chain simulation is carried out in a Jagla-solvent system. The interaction potential between the Jagla-particles and the Jagla-atom of the chain is given $0.80U_A$. Interaction between hard-sphere and Jagla is infinity contact and athermal otherwise. Bond, Angle and Torsion potentials are modeled as below (179-181).

Table 6.1: Energy parameters for Polyalanine model

Bead diameter (σ)	$^{\circ}$A
N_1	3.3
$C_{\alpha,1}$	3.7
C_1	4.0
$CH3_1$	4.4

Bond length (l)	$^{\circ}$A
$N_1 - C_{\alpha,1}$	1.46
$C_{\alpha,1} - C_1$	1.51
$C_1 - N_2$	1.33
$C_{\alpha,1} - CH3_1$	1.53

Bong angle	Degree ($\circ$)
$N - C_{\alpha,1} - C_1$	111.0
$C_{\alpha,1} - C_1 - N_2$	116.0
$C_1 - N_2 - C_{\alpha,2}$	122.0
$N_1 - C_{\alpha,1} - CH3_1$	109.6
$C_1 - C_{\alpha,1} - CH3_1$	110.1

Beads in the protein model are subjected to repulsion due to excluded volume affects, bond pair potential, bond angle potential and torsion potential. The excluded volume is modeled as below

$$u_{ij}(r) = \begin{cases} \infty(r < \sigma_{ij}) \\ 0(r > \sigma_{ij}) \end{cases}$$

Where $u_{ij}(r)$ is the potential between types of atoms i and j when separated by a distance of r and

$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$ where σ is the diameter of the atom.

Bond pair = $K_b(r - \sigma_e)^2$

Bond angle= $K_\theta(\theta - \theta_e)^2$

K_θ and K_b are stiffness constants which in our study is set to 10000. The torsion angle **ω** is held constant at 180°. OPLS torsion potential is used for φ and ψ torsion angles. Monte Carlo simulations were carried out within the range of temperature (T*=0.3 to 1.0) using displacement and rotation moves. We started with a chain in a random configuration. Most of the move sets were being rejected due to the solvent. Hence, to start, solvent molecules were removed. Despite the removal of the solvent molecules the OPLS torsion potential was not able to fold the protein into a helix possibly due to the presence of huge free energy barriers. Subsequently, we analyzed the Ramachandran plot for proteins. Ramachandram plot is a graphical representation which shows the distribution of ψ and φ torsion angles among various secondary structures. The graph shows that in a helical polypeptide chain, the sum of the adjacent ψ and ϕ angles was on average equal to -105°. With this information in hand, we apply an artificial torsion bias potential so that the sum of the adjacent ψ and ϕ angles is equal to -105°. Upon equilibrating, the protein model folds into a helix. Note however, this is in the absence of solvent (**Figure 6.6)**

Future work would involves adding solvent to the system and gradually decreasing the strength of the bias torsion potential to the OPLS potential and subsequently tuning the potential between

the solvent and the proteins. If this can be achieved, then we have shown that a protein chain can be folded into a helix without any specific interaction force (hydrogen bonding).

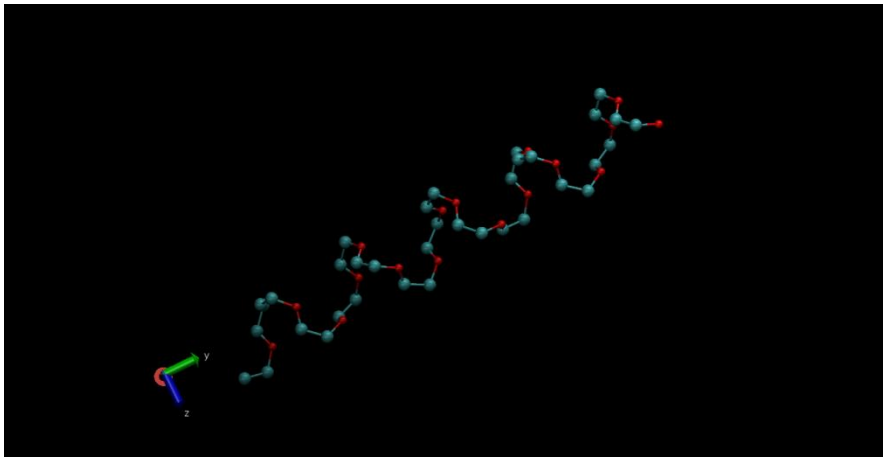


Figure 6.6: Polyalanine in its folded state under biased torsion potential at T*=0.50

Bibliography

1. Zhua Z., A. I. Jewett, S. Kuttimalai, G. Bellesia, S. Gnanakaran, and J.-E. Shea. 2011. Assisted Peptide Folding by Surface Pattern Recognition. Biophysical Journal 100:1306-1315.
2. Kauzmann, W. 1959. SOME FACTORS IN THE INTERPRETATION OF PROTEIN DENATURATION. Advances in Protein Chemistry 14:1-63.
3. Cohn, E. 1919. The changes of haemoglobin as well as protein content in blood serum through muscle work and perspiration. Zeitschrift Fur Biologie 70:366-370.
4. Anson, M. L., and A. E. Mirsky. 1925. On haemochromogen and the relation protein to the properties of the haemoglobin molecule. Journal of Physiology-London 60:50-68.
5. Kamtekar, S., J. M. Schiffer, H. Y. Xiong, J. M. Babik, and M. H. Hecht. 1993. PROTEIN DESIGN BY BINARY PATTERNING OF POLAR AND NONPOLAR AMINO-ACIDS. Science 262:1680-1685.
6. Pauling, L., R. B. Corey, and H. R. Branson. 1951. THE STRUCTURE OF PROTEINS - 2 HYDROGEN-BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN. Proceedings of the National Academy of Sciences of the United States of America 37:205-211.
7. Mirsky, A. E., and L. Pauling. 1936. On the structure of native, denatured, and coagulated proteins. Proceedings of the National Academy of Sciences of the United States of America 22:439-447.
8. Marsh, R. E., R. B. Corey, and L. Pauling. 1955. AN INVESTIGATION OF THE STRUCTURE OF SILK FIBROIN. Biochimica Et Biophysica Acta 16:1-34.
9. Corey, R. B., and L. Pauling. 1953. FUNDAMENTAL DIMENSIONS OF POLYPEPTIDE CHAINS. Proceedings of the Royal Society Series B-Biological Sciences 141:10-20.
10. Varki, A. 1993. BIOLOGICAL ROLES OF OLIGOSACCHARIDES - ALL OF THE THEORIES ARE CORRECT. Glycobiology 3:97-130.
11. Simons, K., and D. Toomre. 2000. Lipid rafts and signal transduction. Nature Reviews Molecular Cell Biology 1:31-39.

12. Ruoslahti, E., and M. D. Pierschbacher. 1987. NEW PERSPECTIVES IN CELL-ADHESION - RGD AND INTEGRINS. Science 238:491-497.
13. Clark, E. A., and J. S. Brugge. 1995. INTEGRINS AND SIGNAL-TRANSDUCTION PATHWAYS - THE ROAD TAKEN. Science 268:233-239.
14. Artavanis-Tsakonas, S., M. D. Rand, and R. J. Lake. 1999. Notch signaling: Cell fate control and signal integration in development. Science 284:770-776.
15. Prime, K. L., and G. M. Whitesides. 1993. ADSORPTION OF PROTEINS ONTO SURFACES CONTAINING END-ATTACHED OLIGO(ETHYLENE OXIDE) - A MODEL SYSTEM USING SELF-ASSEMBLED MONOLAYERS. Journal of the American Chemical Society 115:10714-10721.
16. Jeon, S. I., J. H. Lee, J. D. Andrade, and P. G. Degennes. 1991. PROTEIN SURFACE INTERACTIONS IN THE PRESENCE OF POLYETHYLENE OXIDE .1. SIMPLIFIED THEORY. J. Colloid Interface Sci. 142:149-158.
17. Lampin, M., R. WarocquierClerout, C. Legris, M. Degrange, and M. F. SigotLuizard. 1997. Correlation between substratum roughness and wettability, cell adhesion, and cell migration. Journal of Biomedical Materials Research 36:99-108.
18. Roach, P., D. Farrar, and C. C. Perry. 2005. Interpretation of protein adsorption: Surface-induced conformational changes. Journal of the American Chemical Society 127:8168-8173.
19. Vogler, E. A. 1998. Structure and reactivity of water at biomaterial surfaces. Advances in Colloid and Interface Science 74:69-117.
20. Melzak, K. A., C. S. Sherwood, R. F. B. Turner, and C. A. Haynes. 1996. Driving forces for DNA adsorption to silica in perchlorate solutions. J. Colloid Interface Sci. 181:635-644.
21. Haynes, C. A., E. Sliwinsky, and W. Norde. 1994. STRUCTURAL AND ELECTROSTATIC PROPERTIES OF GLOBULAR-PROTEINS AT A POLYSTYRENE WATER INTERFACE. J. Colloid Interface Sci. 164:394-409.
22. Haynes, C. A., and W. Norde. 1995. STRUCTURES AND STABILITIES OF ADSORBED PROTEINS. J. Colloid Interface Sci. 169:313-328.

23. Arai, T., and W. Norde. 1990. THE BEHAVIOR OF SOME MODEL PROTEINS AT SOLID LIQUID INTERFACES .1. ADSORPTION FROM SINGLE PROTEIN SOLUTIONS. Colloids and Surfaces 51:1-15.
24. Zoungrana, T., G. H. Findenegg, and W. Norde. 1997. Structure, Stability, and Activity of Adsorbed Enzymes. J. Colloid Interface Sci. 190:437-448.
25. Sethuraman, A., and G. Belfort. 2005. Protein structural perturbation and aggregation on homogeneous surfaces. Biophysical Journal 88:1322-1333.
26. Sethuraman, A., G. Vedantham, T. Imoto, T. Przybycien, and G. Belfort. 2004. Protein unfolding at interfaces: Slow dynamics of alpha-helix to beta-sheet transition. Proteins-Structure Function and Bioinformatics 56:669-678.
27. Gray, J. J. 2004. The interaction of proteins with solid surfaces. Current Opinion in Structural Biology 14:110-115.
28. Szleifer, I. 1997. Protein adsorption on surfaces with grafted polymers: A theoretical approach. Biophysical Journal 72:595-612.
29. Hartmann, M. 2005. Ordered mesoporous materials for bioadsorption and biocatalysis. Chemistry of Materials 17:4577-4593.
30. Kim, J., J. W. Grate, and P. Wang. 2006. Nanostructures for enzyme stabilization. Chemical Engineering Science 61:1017-1026.
31. Lai, C. Y., B. G. Trewyn, D. M. Jeftinija, K. Jeftinija, S. Xu, S. Jeftinija, and V. S. Y. Lin. 2003. A mesoporous silica nanosphere-based carrier system with chemically removable CdS nanoparticle caps for stimuli-responsive controlled release of neurotransmitters and drug molecules. Journal of the American Chemical Society 125:4451-4459.
32. Lei, C. H., Y. S. Shin, J. Liu, and E. J. Ackerman. 2002. Entrapping enzyme in a functionalized nanoporous support. Journal of the American Chemical Society 124:11242-11243.
33. Sheldon, R. A. 2007. Enzyme immobilization: The quest for optimum performance. Advanced Synthesis & Catalysis 349:1289-1307.
34. Eggers, D. K., and J. S. Valentine. 2001. Molecular confinement influences protein structure and enhances thermal protein stability. Protein Science 10:250-261.

35. Campanini, B., S. Bologna, F. Cannone, G. Chirico, A. Mozzarelli, and S. Bettati. 2005. Unfolding of Green Fluorescent Protein mut2 in wet nanoporous silica gels. Protein Science 14:1125-1133.
36. Zhou, H. X., and K. A. Dill. 2001. Stabilization of proteins in confined spaces. Biochemistry 40:11289-11293.
37. Zhou, H. X., G. N. Rivas, and A. P. Minton. 2008. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. In Annual Review of Biophysics. 375-397.
38. Minton, A. P. 1992. CONFINEMENT AS A DETERMINANT OF MACROMOLECULAR STRUCTURE AND REACTIVITY. Biophysical Journal 63:1090-1100.
39. Minton, A. P. 1995. CONFINEMENT AS A DETERMINANT OF MACROMOLECULAR STRUCTURE AND REACTIVITY .2. EFFECTS OF WEAKLY ATTRACTIVE INTERACTIONS BETWEEN CONFINED MACROSOLUTES AND CONFINING STRUCTURES. Biophysical Journal 68:1311-1322.
40. Baumketner, A., A. Jewett, and J. E. Shea. 2003. Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape. Journal of Molecular Biology 332:701-713.
41. Betancourt, M. R., and D. Thirumalai. 1999. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. Journal of Molecular Biology 287:627-644.
42. Takagi, F., N. Koga, and S. Takada. 2003. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. Proceedings of the National Academy of Sciences of the United States of America 100:11367-11372.
43. Avnir, D., S. Braun, O. Lev, and M. Ottolenghi. 1994. ENZYMES AND OTHER PROTEINS ENTRAPPED IN SOL-GEL MATERIALS. Chemistry of Materials 6:1605-1614.
44. Martin, C. R. 1994. NANOMATERIALS - A MEMBRANE-BASED SYNTHETIC APPROACH. Science 266:1961-1966.

45. Mateo, C., J. M. Palomo, G. Fernandez-Lorente, J. M. Guisan, and R. Fernandez-Lafuente. 2007. Improvement of enzyme activity, stability and selectivity via immobilization techniques. Enzyme and Microbial Technology 40:1451-1463.
46. Wang, J. 2005. Carbon-nanotube based electrochemical biosensors: A review. Electroanalysis 17:7-14.
47. Brockwell, D. J., E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. 2003. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nature Structural Biology 10:731-737.
48. Hudson, E. P., R. K. Eppler, and D. S. Clark. 2005. Biocatalysis in semi-aqueous and nearly anhydrous conditions. Current Opinion in Biotechnology 16:637-643.
49. Li, S. J., J. Hu, and B. L. Liu. 2004. Use of chemically modified PMMA microspheres for enzyme immobilization. Biosystems 77:25-32.
50. Tran, D. N., and K. J. Balkus. 2011. Perspective of Recent Progress in Immobilization of Enzymes. Acs Catalysis 1:956-968.
51. Berger, B., and T. Leighton. 1998. Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. Journal of Computational Biology 5:27-40.
52. Dill, K. A., S. Bromberg, K. Z. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. 1995. PRINCIPLES OF PROTEIN-FOLDING - A PERSPECTIVE FROM SIMPLE EXACT MODELS. Protein Science 4:561-602.
53. Yue, K., K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Shakhnovich, and K. A. Dill. 1995. A TEST OF LATTICE PROTEIN-FOLDING ALGORITHMS. Proceedings of the National Academy of Sciences of the United States of America 92:325-329.
54. Siepmann, J. I. 1999. An introduction to the Monte Carlo method for particle simulations. Monte Carlo Methods in Chemical Physics 105:1-12.
55. Cheng, S.-R. 2009. Highly nonlinear model in finance and convergence of Monte Carlo simulations. Journal of Mathematical Analysis and Applications 353:531-543.
56. Lemieux, C. 2004. Randomized quasi-Monte Carlo: A tool for improving the efficiency of simulations in finance.
57. Liberati, N. B., and E. Platen. 2004. On the efficiency of simplified weak Taylor schemes for Monte Carlo simulation in finance. In Computational Science - Iccs 2004,

Proceedings. M. Bubak, G. DickVanAlbada, P. M. A. Sloot, and J. J. Dongarra, editors. 771-778.

58. Carley, D. D. 1971. RADIAL DISTRIBUTION FUNCTIONS FOR A BINARY MIXTURE OF GAUSSIAN MOLECULES. Journal of Chemical Physics 55:3426-&.

59. Eichinger, B. E., and C. D. Howell. 1977. DISTRIBUTION FUNCTIONS FOR GAUSSIAN MOLECULES. Abstracts of Papers of the American Chemical Society 174:117-117.

60. Tsapline, B., M. Roux, and M. Cornille. 1970. STUDY OF DISTRIBUTION OF MOMENTS IN ATOMS AND MOLECULES. CASE OF WAVE FUNCTIONS DEVELOPED ON BASE OF GAUSSIAN FUNCTIONS. Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences Serie B 270:865-&.

61. Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. EQUATION OF STATE CALCULATIONS BY FAST COMPUTING MACHINES. Journal of Chemical Physics 21:1087-1092.

62. Chin, S. A. 1990. QUADRATIC DIFFUSION MONTE-CARLO ALGORITHMS FOR SOLVING ATOMIC MANY-BODY PROBLEMS. Physical Review A 42:6991-7005.

63. Melchers, R. E. 1990. SEARCH-BASED IMPORTANCE SAMPLING. Structural Safety 9:117-128.

64. Paine, G. H., and H. A. Scheraga. 1985. PREDICTION OF THE NATIVE CONFORMATION OF A POLYPEPTIDE BY A STATISTICAL-MECHANICAL PROCEDURE .1. BACKBONE STRUCTURE OF ENKEPHALIN. Biopolymers 24:1391-1436.

65. Hastings, W. K. 1970. MONTE-CARLO SAMPLING METHODS USING MARKOV CHAINS AND THEIR APPLICATIONS. Biometrika 57:97-&.

66. Haario, H., E. Saksman, and J. Tamminen. 2001. An adaptive Metropolis algorithm. Bernoulli 7:223-242.

67. Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. Journal of the Physical Society of Japan 65:1604-1608.

68. Wood, W. W., and J. D. Jacobson. 1957. PRELIMINARY RESULTS FROM A RECALCULATION OF THE MONTE CARLO EQUATION OF STATE OF HARD SPHERES. Journal of Chemical Physics 27:1207-1208.
69. Adams, D. J. 1975. GRAND CANONICAL ENSEMBLE MONTE-CARLO FOR A LENNARD-JONES FLUID. Molecular Physics 29:307-311.
70. Chesnut, D. A., and Z. W. Salsburg. 1963. MONTE CARLO PROCEDURE FOR STATISTICAL MECHANICAL CALCULATIONS IN A GRAND CANONICAL ENSEMBLE OF LATTICE SYSTEMS. Journal of Chemical Physics 38:2861-&.
71. Rowley, L. A., D. Nicholson, and N. G. Parsonage. 1975. MONTE-CARLO GRAND CANONICAL ENSEMBLE CALCULATION IN A GAS-LIQUID TRANSITION REGION FOR 12-6 ARGON. Journal of Computational Physics 17:401-414.
72. Manka, A., W. Nowicki, and G. Nowicka. 2013. Monte Carlo simulations of a polymer chain conformation. The effectiveness of local moves algorithms and estimation of entropy. Journal of Molecular Modeling 19:3659-3670.
73. Stockmayer, W. H. 1944. Theory of molecular size distribution and gel formation in branched polymers II General cross linking. Journal of Chemical Physics 12:125-131.
74. Zimm, B. H., and W. H. Stockmayer. 1949. THE DIMENSIONS OF CHAIN MOLECULES CONTAINING BRANCHES AND RINGS. Journal of Chemical Physics 17:1301-1314.
75. Chodera, J. D., W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill. 2007. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. Journal of Chemical Theory and Computation 3:26-41.
76. Earl, D. J., and M. W. Deem. 2005. Parallel tempering: Theory, applications, and new perspectives. Physical Chemistry Chemical Physics 7:3910-3916.
77. Hansmann, U. H. E. 1997. Parallel tempering algorithm for conformational studies of biological molecules. Chemical Physics Letters 281:140-150.
78. Liu, P., B. Kim, R. A. Friesner, and B. J. Berne. 2005. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proceedings of the National Academy of Sciences of the United States of America 102:13749-13754.

79. Landau, D. P., S. H. Tsai, and M. Exler. 2004. A new approach to Monte Carlo simulations in statistical physics: Wang-Landau sampling. American Journal of Physics 72:1294-1302.
80. Shell, M. S., P. G. Debenedetti, and A. Z. Panagiotopoulos. 2002. Generalization of the Wang-Landau method for off-lattice simulations. Physical Review E 66.
81. Shell, M. S., P. G. Debenedetti, and A. Z. Panagiotopoulos. 2003. An improved Monte Carlo method for direct calculation of the density of states. Journal of Chemical Physics 119:9406-9411.
82. Trebst, S., D. A. Huse, and M. Troyer. 2004. Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations. Physical Review E 70.
83. Zhou, C. G., and R. N. Bhatt. 2005. Understanding and improving the Wang-Landau algorithm. Physical Review E 72.
84. Andrews, T. M., and J. R. Tata. 1971. PROTEIN SYNTHESIS BY MEMBRANE-BOUND AND FREE RIBOSOMES OF SECRETORY AND NON-SECRETORY TISSUES. Biochemical Journal 121:683-&.
85. Hartl, F. U., A. Bracher, and M. Hayer-Hartl. 2011. Molecular chaperones in protein folding and proteostasis. Nature 475:324-332.
86. Jin, W., and J. D. Brennan. 2002. Properties and applications of proteins encapsulated within sol-gel derived materials. Analytica Chimica Acta 461:1-36.
87. Martin, J., and F. U. Hartl. 1997. Chaperone-assisted protein folding. Current Opinion in Structural Biology 7:41-52.
88. Martin, J., T. Langer, R. Boteva, A. Schramel, A. L. Horwich, and F. U. Hartl. 1991. CHAPERONIN-MEDIATED PROTEIN FOLDING AT THE SURFACE OF GROEL THROUGH A MOLTEN GLOBULE-LIKE INTERMEDIATE. Nature 352:36-42.
89. Ellis, R. J. 2001. Macromolecular crowding: obvious but underappreciated. Trends in Biochemical Sciences 26:597-604.
90. Ellis, R. J., and A. P. Minton. 2003. Cell biology - Join the crowd. Nature 425:27-28.
91. Zimmerman, S. B., and A. P. Minton. 1993. MACROMOLECULAR CROWDING - BIOCHEMICAL, BIOPHYSICAL, AND PHYSIOLOGICAL CONSEQUENCES. Annual Review of Biophysics and Biomolecular Structure 22:27-65.

92. Conway, K. A., S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, and P. T. Lansbury. 2000. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. Proceedings of the National Academy of Sciences of the United States of America 97:571-576.

93. Kelly, J. W. 1998. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Current Opinion in Structural Biology 8:101-106.

94. Kelly, J. W. 2002. Towards an understanding of amyloidogenesis. Nature Structural Biology 9:323-325.

95. Kirschner, D. A., C. Abraham, and D. J. Selkoe. 1986. X-RAY-DIFFRACTION FROM INTRANEURONAL PAIRED HELICAL FILAMENTS AND EXTRANEURONAL AMYLOID FIBERS IN ALZHEIMER-DISEASE INDICATES CROSS-BETA CONFORMATION. Proceedings of the National Academy of Sciences of the United States of America 83:503-507.

96. Koo, E. H., P. T. Lansbury, and J. W. Kelly. 1999. Amyloid diseases: Abnormal protein aggregation in neurodegeneration. Proceedings of the National Academy of Sciences of the United States of America 96:9989-9990.

97. Corrales, F. J., and A. R. Fersht. 1996. Toward a mechanism for GroEL center dot GroES chaperone activity: An ATPase-gated and -pulsed folding and annealing cage. Proceedings of the National Academy of Sciences of the United States of America 93:4509-4512.

98. Jewett, A. I., and J. E. Shea. 2006. Folding on the chaperone: Yield enhancement through loose binding. Journal of Molecular Biology 363:945-957.

99. Wang, G., and S. T. Yau. 2007. Spatial confinement induced enzyme stability for bioelectronic applications. Journal of Physical Chemistry C 111:11921-11926.

100. Wang, Y. J., and F. Caruso. 2005. Mesoporous silica spheres as supports for enzyme immobilization and encapsulation. Chemistry of Materials 17:953-961.

101. Rathore, N., T. A. Knotts, and J. J. de Pablo. 2006. Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations. Biophysical Journal 90:1767-1773.

102. Andrade, J. D., and V. Hlady. 1986. PROTEIN ADSORPTION AND MATERIALS BIOCOMPATIBILITY - A TUTORIAL REVIEW AND SUGGESTED HYPOTHESES. Advances in Polymer Science 79:1-63.
103. Felsovalyi, F., P. Mangiagalli, C. Bureau, S. K. Kumar, and S. Banta. 2011. Reversibility of the Adsorption of Lysozyme on Silica. Langmuir 27:11873-11882.
104. Haynes, C. A., and W. Norde. 1994. Globular proteins at solid/liquid interfaces. Colloids and Surfaces B: Biointerfaces 2:517-566.
105. Norde, W., and J. P. Favier. 1992. Structure of adsorbed and desorbed proteins. Colloids and Surfaces 64:87-93.
106. Raffaini, G., and F. Ganazzoli. 2010. Protein Adsorption on a Hydrophobic Surface: A Molecular Dynamics Study of Lysozyme on Graphite. Langmuir 26:5679-5689.
107. Vermeer, A. W. P., C. E. Giacomelli, and W. Norde. 2001. Adsorption of IgG onto hydrophobic teflon. Differences between the Fab and Fc domains. Biochimica et Biophysica Acta (BBA) - General Subjects 1526:61-69.
108. Sharma, S., B. J. Berne, and S. K. Kumar. 2010. Thermal and Structural Stability of Adsorbed Proteins. Biophysical Journal 99:1157-1165.
109. Sang, L. C., and M. O. Coppens. 2011. Effects of surface curvature and surface chemistry on the structure and activity of proteins adsorbed in nanopores. Physical Chemistry Chemical Physics 13:6689-6698.
110. Dill, K. A. 1985. THEORY FOR THE FOLDING AND STABILITY OF GLOBULAR-PROTEINS. Biochemistry 24:1501-1509.
111. Yue, K., and K. A. Dill. 1995. FORCES OF TERTIARY STRUCTURAL ORGANIZATION IN GLOBULAR-PROTEINS. Proceedings of the National Academy of Sciences of the United States of America 92:146-150.
112. Crescenzi, P., D. Goldman, C. Papadimitriou, A. Piccolboni, and M. Yannakakis. 1998. On the complexity of protein folding. Journal of Computational Biology 5:423-465.
113. Siepmann, J. I., and D. Frenkel. 1992. CONFIGURATIONAL BIAS MONTE-CARLO - A NEW SAMPLING SCHEME FOR FLEXIBLE CHAINS. Molecular Physics 75:59-70.

114. Ferrenberg, A. M., and R. H. Swendsen. 1989. OPTIMIZED MONTE-CARLO DATA-ANALYSIS. Physical Review Letters 63:1195-1198.
115. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD. Journal of Computational Chemistry 13:1011-1021.
116. Friedel, M., D. J. Sheeler, and J. E. Shea. 2003. Effects of confinement and crowding on the thermodynamics and kinetics of folding of a minimalist beta-barrel protein. Journal of Chemical Physics 118:8106-8113.
117. Zhou, H. X. 2008. Effect of mixed macromolecular crowding agents on protein folding. Proteins-Structure Function and Bioinformatics 72:1109-1113.
118. Kim, J., H. F. Jia, and P. Wang. 2006. Challenges in biocatalysis for enzyme-based biofuel cells. Biotechnology Advances 24:296-308.
119. Luckarift, H. R., J. C. Spain, R. R. Naik, and M. O. Stone. 2004. Enzyme immobilization in a biomimetic silica support. Nature Biotechnology 22:211-213.
120. Verma, M. L., C. J. Barrow, and M. Puri. 2013. Nanobiotechnology as a novel paradigm for enzyme immobilisation and stabilisation with potential applications in biodiesel production. Appl. Microbiol. Biotechnol. 97:23-39.
121. Wang, S. S., and W. R. Vieth. 1973. COLLAGEN-ENZYME COMPLEX MEMBRANES AND THEIR PERFORMANCE IN BIOCATALYTIC MODULES. Biotechnology and Bioengineering 15:93-115.
122. Asuri, P., S. S. Karajanagi, H. C. Yang, T. J. Yim, R. S. Kane, and J. S. Dordick. 2006. Increasing protein stability through control of the nanoscale environment. Langmuir 22:5833-5836.
123. Litt, J., C. Padala, P. Asuri, S. Vutukuru, K. Athmakuri, S. Kumar, J. Dordick, and R. S. Kane. 2009. Enhancing Protein Stability by Adsorption onto Raftlike Lipid Domains. Journal of the American Chemical Society 131:7107-7111.
124. Hall, C. K. 2008. Thermodynamic and kinetic origins of Alzheimer's and related diseases: a chemical engineer's perspective. Aiche Journal 54:1956-1962.

125. Nguyen, H. D., and C. K. Hall. 2004. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proceedings of the National Academy of Sciences of the United States of America 101:16180-16185.
126. Liberek, K., A. Lewandowska, and S. Zietkiewicz. 2008. Chaperones in control of protein disaggregation. Embo Journal 27:328-335.
127. Rosenzweig, R., S. Moradi, A. Zarrine-Afsar, J. R. Glover, and L. E. Kay. 2013. Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction. Science 339:1080-1083.
128. Radhakrishna, M., J. Grimaldi, G. Belfort, and S. K. Kumar. 2013. Stability of Proteins Inside a Hydrophobic Cavity. Langmuir.
129. Chiti, F., and C. M. Dobson. 2009. Amyloid formation by globular proteins under native conditions. Nature Chemical Biology 5:15-22.
130. Patel, A. J., P. Varilly, S. N. Jamadagni, H. Acharya, S. Garde, and D. Chandler. 2011. Extended surfaces modulate hydrophobic interactions of neighboring solutes. Proceedings of the National Academy of Sciences of the United States of America 108:17678-17683.
131. Ueda, Y., H. Taketomi, and N. Gō. 1978. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. Biopolymers 17:1531-1548.
132. Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. Proceedings of the National Academy of Sciences of the United States of America 100:13898-13903.
133. Guo, Z. Y., D. Thirumalai, and J. D. Honeycutt. 1992. FOLDING KINETICS OF PROTEINS - A MODEL STUDY. Journal of Chemical Physics 97:525-535.
134. Torrie, G. M., and J. P. Valleau. 1977. NON-PHYSICAL SAMPLING DISTRIBUTIONS IN MONTE-CARLO FREE-ENERGY ESTIMATION - UMBRELLA SAMPLING. Journal of Computational Physics 23:187-199.
135. Wust, T., and D. P. Landau. 2009. Versatile Approach to Access the Low Temperature Thermodynamics of Lattice Polymers and Proteins. Physical Review Letters 102.

136. Deutsch, J. M. 1997. Long range moves for high density polymer simulations. Journal of Chemical Physics 106:8849-8854.
137. Jain, T. S., and J. J. de Pablo. 2002. A biased Monte Carlo technique for calculation of the density of states of polymer films. Journal of Chemical Physics 116:7238-7243.
138. Rathore, N., and J. J. de Pablo. 2002. Monte Carlo simulation of proteins through a random walk in energy space. Journal of Chemical Physics 116:7225-7230.
139. Rathore, N., T. A. Knotts, and J. J. de Pablo. 2003. Density of states simulations of proteins. Journal of Chemical Physics 118:4285-4290.
140. Rathore, N., T. A. Knotts, and J. J. de Pablo. 2003. Configurational temperature density of states simulations of proteins. Biophysical Journal 85:3963-3968.
141. Castells, V., and P. R. Van Tassel. 2005. Conformational transition free energy profiles of an adsorbed, lattice model protein by multicanonical Monte Carlo simulation. Journal of Chemical Physics 122.
142. Li, Y. W., T. Wust, and D. P. Landau. 2011. Monte Carlo simulations of the HP model (the "Ising model" of protein folding). Computer Physics Communications 182:1896-1899.
143. Anand, G., S. Sharma, A. K. Dutta, S. K. Kumar, and G. Belfort. 2010. Conformational Transitions of Adsorbed Proteins on Surfaces of Varying Polarity. Langmuir 26:10803-10811.
144. Anand, G., S. Sharma, S. K. Kumar, and G. Belfort. 2009. Stability of Tethered Proteins. Langmuir 25:4998-5005.
145. Anand, G., S. N. Jamadagni, S. Garde, and G. Belfort. 2010. Self assembly of TMAO at hydrophobic interfaces and its effect on protein adsorption: Insights from experiments and simulations. Abstracts of Papers of the American Chemical Society 240.
146. Cellmer, T., D. Bratko, and H. Blanch. 2003. Langevin dynamics simulations of protein aggregation. Biophysical Journal 84:41A-42A.
147. Shakhnovich, E. I., and A. M. Gutin. 1993. A NEW APPROACH TO THE DESIGN OF STABLE PROTEINS. Protein Engineering 6:793-800.

148. Shakhnovich, E. I., and A. M. Gutin. 1993. ENGINEERING OF STABLE AND FAST-FOLDING SEQUENCES OF MODEL PROTEINS. Proceedings of the National Academy of Sciences of the United States of America 90:7195-7199.
149. Bellesia, G., A. I. Jewett, and J. E. Shea. 2010. Sequence periodicity and secondary structure propensity in model proteins. Protein Science 19:141-154.
150. Jagla, E. A. 1999. Core-softened potentials and the anomalous properties of water. Journal of Chemical Physics 111:8980-8986.
151. Jagla, E. A. 2001. Low-temperature behavior of core-softened models: Water and silica behavior. Physical Review E 63:art. no.-061509.
152. Kabsch, W., and C. Sander. 1983. DICTIONARY OF PROTEIN SECONDARY STRUCTURE - PATTERN-RECOGNITION OF HYDROGEN-BONDED AND GEOMETRICAL FEATURES. Biopolymers 22:2577-2637.
153. Kitagawa, S., R. Kitaura, and S. Noro. 2004. Functional porous coordination polymers. Angewandte Chemie-International Edition 43:2334-2375.
154. Mark, P., and L. Nilsson. 2001. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. Journal of Physical Chemistry A 105:9954-9960.
155. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B 102:3586-3616.
156. Hall, C. K., and G. Stell. 1973. PHASE-TRANSITIONS IN 2-DIMENSIONAL LATTICE GASES OF HARD-CORE MOLECULES WITH LONG-RANGE ATTRACTIONS. Physical Review A 7:1679-1689.
157. Stell, G., and P. C. Hemmer. 1972. PHASE-TRANSITIONS DUE TO SOFTNESS OF POTENTIAL CORE. Journal of Chemical Physics 56:4274-&.
158. Hemmer, P. C., and G. Stell. 1970. FLUIDS WITH SEVERAL PHASE TRANSITIONS. Physical Review Letters 24:1284-&.

159. Xu, L., S. V. Buldyrev, C. A. Angell, and H. E. Stanley. 2006. Thermodynamics and dynamics of the two-scale spherically symmetric Jagla ramp model of anomalous liquids. Physical Review E 74.

160. Jagla, E. A. 2001. Liquid-liquid equilibrium for monodisperse spherical particles. Physical Review E 63.

161. Jagla, E. A. 1998. Phase behavior of a system of particles with core collapse. Physical Review E 58:1478-1486.

162. Buldyrev, S. V., P. Kumar, P. G. Debenedetti, P. J. Rossky, and H. E. Stanley. 2007. Water-like solvation thermodynamics in a spherically symmetric solvent model with two characteristic lengths. Proceedings of the National Academy of Sciences of the United States of America 104:20177-20182.

163. Xu, L., N. Giovambattista, S. V. Buldyrev, P. G. Debenedetti, and H. E. Stanley. 2011. Waterlike glass polyamorphism in a monoatomic isotropic Jagla model. Journal of Chemical Physics 134.

164. Stone, M. T., P. J. In 't Veld, Y. Lu, and I. C. Sanchez. 2002. Hydrophobic/hydrophilic solvation: inferences from Monte Carlo simulations and experiments. Molecular Physics 100:2773-2792.

165. Maiti, M., S. Weiner, S. V. Buldyrev, H. E. Stanley, and S. Sastry. 2012. Potential of mean force between hydrophobic solutes in the Jagla model of water and implications for cold denaturation of proteins. Journal of Chemical Physics 136.

166. Patel, H. A., E. B. Nauman, and S. Garde. 2003. Molecular structure and hydrophobic solvation thermodynamics at an octane-water interface. Journal of Chemical Physics 119:9199-9206.

167. Garde, S. 2011. Hydrophobicity of proteins and interfaces: Insights from density fluctuations. Abstracts of Papers of the American Chemical Society 241.

168. Matysiak, S., P. G. Debenedetti, and P. J. Rossky. 2012. Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation. Journal of Physical Chemistry B 116:8095-8104.

169. Zhang, D., Z. Yang, X. Wen, Z. Xiang, L. He, S. Ran, and L. Zhang. 2011. Helical Conformations of Semiflexible Polymers Confined between Two Concentric Cylinders. Journal of Physical Chemistry B 115:14333-14340.

170. Haliloglu, T., and I. Bahar. 1998. Coarse-grained simulations of conformational dynamics of proteins: Application to apomyoglobin. Proteins-Structure Function and Genetics 31:271-281.

171. Zimm, B. H., and J. K. Bragg. 1959. THEORY OF THE PHASE TRANSITION BETWEEN HELIX AND RANDOM COIL IN POLYPEPTIDE CHAINS. Journal of Chemical Physics 31:526-535.

172. Ramakrishnan, V., R. Ranbhor, and S. Durani. 2005. Simulated folding in polypeptides of diversified molecular tacticity: Implications for protein folding and de novo design. Biopolymers 78:96-105.

173. Takada, S., Z. Luthey-Schulten, and P. G. Wolynes. 1999. Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. Journal of Chemical Physics 110:11616-11629.

174. Nguyen, H. D., A. J. Marchut, and C. K. Hall. 2004. Solvent effects on the conformational transition of a model polyalanine peptide. Protein Science 13:2909-2924.

175. Sorin, E. J., Y. M. Rhee, M. R. Shirts, and V. S. Pande. 2006. The solvation interface is a determining factor in peptide conformational preferences. Journal of Molecular Biology 356:248-256.

176. Jacchieri, S. G. 1992. AN EXAMINATION OF THE BASIC HYPOTHESIS OF ZIMM-BRAGG THEORY BASED ON ENERGY-DISTRIBUTIONS OF PEPTIDE CHAINS. International Journal of Quantum Chemistry:255-272.

177. Scholtz, J. M., and R. L. Baldwin. 1992. THE MECHANISM OF ALPHA-HELIX FORMATION BY PEPTIDES. Annual Review of Biophysics and Biomolecular Structure 21:95-118.

178. Sharma, S., S. K. Kumar, S. V. Buldyrev, P. G. Debenedetti, P. J. Rossky, and H. E. Stanley. 2013. A Coarse-Grained Protein Model in a Water-like Solvent. Scientific Reports 3.

179. Hall, C. K., and A. V. Smith. 2001. Molecular dynamics simulations of multiprotein systems. Biophysical Journal 80:557A-557A.
180. Smith, A. V., and C. K. Hall. 2001. alpha-helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. Proteins-Structure Function and Genetics 44:344-360.
181. Smith, A. V., and C. K. Hall. 2001. Assembly of a tetrameric alpha-helical bundle: Computer simulations on an intermediate-resolution protein model. Proteins-Structure Function and Genetics 44:376-391.